Me, Myself and I: Reflections on Self-Consciousness and Authority

ME, MYSELF, AND I: REFLECTIONS ON SELF-CONSCIOUSNESS AND AUTHORITY

A Dissertation Presented

by

JONATHAN SCOTT ROSEN

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements of

DOCTOR OF PHILOSOPHY

September 2017

Philosophy

ME, MYSELF, AND I: REFLECTIONS ON SELF-CONSCIOUSNESS AND AUTHORITY

A Dissertation Presented
by
JONATHAN SCOTT ROSEN

Approved as to style and content by:

Hilary Kornblith, Chair

Peter Graham, Member

Vanessa de Harven, Member

Andy Rotman, Member

Joseph Levine, Department Head
Philosophy

DEDICATION

To Leo and Benny

ACKNOWLEDGEMENTS

When, ten years ago, I applied to the UMass, Amherst philosophy graduate program, I originally asked to be admitted into the Masters Program—this, because the prospect of pursuing a PhD as a 40-year-old, with young kids, was unimaginable. By "unimaginable," I don't mean that I literally could not imagine it. I *could* imagine it—just like I could imagine running a four-minute mile or climbing Mt. Everest. The fact was that I suspected the difficulties involved—the sheer quantity and complexity of the work expected of me—would easily overwhelm me. To pursue a PhD would be "crazy".

What I did not realize at the time was that my estimate of the quantity and difficulty of the work involved in earning a PhD would fall dramatically short of the mark, and that the pursuit—which, after my first year, I decided to undertake—would *not only* drive me "crazy," but would do so on a regular basis, reducing me to states of panic and desolation. Part of this difficulty owed to the fact that, given my artistic/literary background, I was not accustomed to the spartan, scalpel-wrought style of so much of analytic philosophy. Part of it lay in the treachery of the philosophical problems themselves. Part of it lay in the demands on clear philosophical writing; and, further, in what I took to be an additional demand—that is, to produce writing that was accessible and engaging.

Yet there was a further difficulty that threatened to undermine the academic philosophical enterprise as a whole. That is: How can one reasonably hope to shed additional light on problems that have troubled great thinkers for hundreds or

thousands of years? And, relatedly: Given the superabundance of scholarship on such problems—scholarship produced by minds brighter and better educated than my own—when could one justifiably conclude that one had read enough, or thought enough, such that one could weigh in on the matter at hand? What is it that renders one capable of such "forward motion"?

Speaking about the creative process, the novelist Flannery O'Conner has written: "I have always insisted that there is a fine grain of stupidity required in the fiction writer." I believe the same holds true for the philosophical enterprise. That is, one might easily imagine a hyper-conscientious student who would *never* make any progress; who would never comfortably conclude that he had read enough, or thought enough; who would never have the audacity to venture his "own" ideas. As such, I suggest that every forward step must involve something arbitrary, blind, or, as O'Connor put it, stupid. And *one knows this*. And yet, because one must get work done by deadline, one must *blind oneself to this understanding*. And it is this conflict—on one hand, (1) a neglect necessitated by the demand of forward academic motion; and (2) the need for rigorous, responsible circumspection that is the hallmark of analytic philosophy—that produces the further difficulty, or madness, that I referred to above.

Operating against this background, I realized that I would need to find a set of thinkers who engaged in philosophy the way I naturally did; who articulated and explored problems in a style that was congenial to mine. Just as importantly, I would need to find professors who were sympathetic to my particular bent, who could appreciate my own take on such problems and who appreciated the

peculiarities of the philosophical enterprise, as I saw them. And here I was very fortunate. The thinkers are all on display in the dissertation. The professors of particular assistance were Ernesto Garcia, who helped me develop an understanding of Kant and Hume; Lynne Baker who introduced me to Wittgenstein and shone much light on the first person perspective (and served as an understanding and helpful advisor); and Gary Matthews and Vanessa de Harven who deepened my understanding and appreciation of Plato and Aristotle. Of additional help were Andy Rotman and Nalini Bhushan with whom I have worked at Smith College and who brought a spirit of fun to philosophy, and Pete Graham who served on my dissertation committee.

Yet my academic survival owes particularly to the guidance of Hilary Kornblith, who, with Lynne and Gary, wrote one of my initial letters of recommendation, and whose manner of teaching, and straightforward style of writing, served as an immense inspiration. More than this, Hilary always made himself available in his comfortable office, responded with uncanny promptness to my e-mails, and offered very helpful feedback and encouragement when I was struggling. I am immensely grateful for his steadfast assistance, encouragement and friendship.

Beyond these generous faculty members, I owe a huge debt of gratitude to my friend Matt Wohl, with whom I discussed all my ideas and who read nearly all my papers and many of the essays assigned for class. Our extended conversations helped clarify much of my thinking, and, as a result, helped me convey my thoughts in writing. Thanks also to Dan Seitz for encouragement and advice,

Walter Krudop for producing the dissertation diagrams, Nik Contis for thesis title suggestions, and Rachel Robison for collegiality, humor, and friendship.

Finally, an especially deep thank-you to my sons Leo and Benny, to the support of my parents, Mel and Ellen, my parents-in-law Michel and Sibylle Baier, and, of course, to my lovely wife, Julia, who softened so many blows, and gave me so many reasons to persist in this "crazy" philosophical endeavor, which, all told, has turned out to be the most intellectually challenging and gratifying experience of my life.

ABSTRACT

ME, MYSELF AND I: REFLECTIONS ON SELF-CONSCIOUSNESS AND AUTHORITY

SEPTEMBER 2017

JONATHAN SCOTT ROSEN

B.A. HAMPSHIRE COLLEGE

Ph.D. UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Hilary Kornblith

The Rationalist conception of the self identifies the subject, the "I", as a "captain" wielding autonomous rational authority over his subservient attitudes and behaviors—his "crew". I argue that such a conception of the self is metaphysically untenable and that its practical and ethical ramifications are unattractive. In its place I recommend an alternative, Holistic, "Crew of Captains" conception of the self, and explain its metaphysical, practical and ethical advantages.

PREFACE

Most persons who have written about the affects and man's conduct of life seem to discuss, not the natural things which follow the common laws of nature, but things which are outside her. They seem indeed to consider man in nature as a kingdom within a kingdom. For they believe that man disturbs rather than follows her order; that he has an absolute power over his own actions; and that he is altogether self-determined.

--Spinoza, *Ethics* III, Preface.

TABLE OF CONTENTS

Page

CHAPTER 1

INTRODUCTION

1.1 Introduction

Like non-human animals, human beings take a great interest in the world. Like them, we seek nourishment, shelter, and reproductive partners. Unlike non-human animals, we also take a great deal of interest in ourselves. We wish to "understand ourselves," "take ourselves seriously," "take charge of ourselves," "find ourselves," "improve ourselves," etc. Indeed, a glance at the table of contents of a standard "self-help" book or a typical "wellness" retreat course catalogue will reveal a plethora of approaches one may take toward the understanding, development, cultivation and improvement of one's "self." As Charles Guignon has put it: "The central issue for modernity is *autonomy*, or self-direction, being the captain of your own ship. What we hope to achieve in life is not honor as that was traditionally conceived, but rather the *dignity* that arises from being a bounded, masterful, autonomous self."[1]

What might it mean, to be a "bounded, masterful, autonomous self," to play the role of "captain" with respect to our "ships,"[2] and on what basis do we believe we are capable of becoming so? With respect to the first question, I suggest that

[1] Charles Guignon, *On Being Authentic* (Routledge, 2004), p. 150, emphasis his.

[2] Note that I will be employing this "Captain" analogy of self-governance throughout my discussion. However, instead of referring to the self as the "Captain-of-Ship," I will refer to it as the "Captain-of-Crew." I believe the latter expression better captures the putative relationship between the self *qua* captain and the captain's subservient parts or "crew members." It will also better position me to offer an alternative model of the self that I shall call the "Crew-of-Captains" model.

we become bounded, masterful and autonomous—or at least, we believe we have become so—when we have satisfied at least three conditions: (1) When we have achieved a sufficient understanding of ourselves (our thoughts, feelings and behavior do not surprise or mystify us and we can reliably predict how we will behave under possible circumstances); (2) When we have assumed ownership of or taken command of our actions (our actions are not the result of "external" forces pushing us around but are in fact *up to* us); and, (3) When we have organized or fashioned ourselves into the type of people we wish to be (we cultivate qualities we approve of and we reduce or modify our undesired attributes), or, at least, when we are in the process of doing so.

With respect to the second question, I suggest that our confidence in our ability to become bounded, masterful, autonomous selves derives from our faith in the Rationalist Story[3]. The Rationalist Story may be broken into four parts that can be summarized as follows. Part 1: We human beings have the capacity to take a "backward step," that is, to assume a self-conscious reflective stance. Part 2: The assumption of the self-conscious reflective stance divides the self into what William James calls a "duplex"; into, on one hand, an empirical, factical, observable element (what James called the "me"); and into, on the other hand, a subjective element that can observe and attend to the empirical element (what James called the "I").[4] Part 3: Not only does self-consciousness divide the self

[3] The employment of the term "Rationalist" is not meant to express any commitments along rationalist/empiricist lines. Rather, the term is meant to capture the importance of the exercise of *rational agency* stressed by proponents of this account. See a similar usage in Cassam (2014) and Gertler (2011).

[4] See William James, *The Principles of Psychology, Vol. 1*, (Dover Publications: NY, 1918), pp. 291-401, and William James, *Psychology, The Briefer Course,*

into these two parts, but, in doing so, self-consciousness opens up a space, or a *reflective distance* between the "I" and the "me".[5] Lastly, Part 4: In virtue of attaining such a reflective distance, the subjective self inherits a repertoire of powers it can exercise with respect to the empirical self.

What kind of powers can the subjective "I" exercise with respect to the empirical "me," in virtue of having achieved such reflective distance?[6] I refer again to the three powers mentioned above. First, an ability to acquire self-knowledge. When my attention is primarily focused out on the external world, I

(Dover Publications: NY, 2001), pp. 42-83. Of course the notion of the "Duplex Self" that arises from self-consciousness does not originate with James. See, for example, Kant's notion of the empirical and intelligible (or *noumenal*) selves (e.g., Immanuel Kant, *Critique of Practical Reason*, (ed.) Mary Gregor, (intro.) Andrews Reath (Cambridge, 1997), pp. 81-83/ 5:97-99. I will also address Sartre's empirical/factical "being -in-itself" (*etre-en-soi*) versus his transcendental "being-for-itself" (*etre-pour-soi*). Jean Paul Sartre, *Being and Nothingness* (trans.) Hazel Barnes, (Washington Square Press, 1993).

[5] Christine Korsgaard appears to take credit for the expression "reflective distance". See Korsgaard: "Once the space of awareness—of reflective distance, as I like to call it—opens up between the potential ground of a belief and the belief itself, or between the potential ground of an action and the action itself, we must step across that distance with some awareness that we are doing so…" Christine Korsgaard, *The Constitution of Agency*, *Essays on Practical Reason and Moral Psychology* (Oxford, 2008), p. 4; and see also Korsgaard (2009, p. 116). But see also Velleman: "Most importantly, though, consciousness just seems to open a gulf between subject and object, even when its object is the subject himself. Consciousness seems to have the structure of vision, requiring its object to stand across from the viewer—to occupy the position of *Gegenstand* [object or thing]." J. David Velleman, "The Way of the Wanton," from *Practical Identity and Narrative Agency*, (eds.) Kim Watkins and Catriona, (Routledge, 2008, p. 1790. And see Nagel: "This step back, this opening of a slight space between inclination and decision, is he condition that permits the operation of reason…" Thomas Nagel, *The Last Word*, (Oxford, 1997), p. 109. I will examine these, and additional characterizations of such "space," as we proceed.

[6] The terms I (no quotation marks), "I", "me," self, and person will be put to various uses over the course of this thesis. In Section 2.3, I will offer a more concise explication of these terms. For the time being, I ask the reader to recognize the term I, without quotation marks, as signifying the "human being" I am (the public, human, individual named Jon Rosen). When I am referring to the Jamesian reflective "I," I will use quotation marks.

do not self-consciously reflect on, or evaluate, myself. Once I take the backward step, however, I can view myself from a distance, just as I view objects in the world, and—so the Rationalist tells us—I can thereby acquire a good deal of knowledge about my empirical self, just as I can acquire a good deal of knowledge about objects in the world. Second, the capacity for "deep"[7] agency or autonomy. When I am unreflectively engaged in worldly affairs and am not considering my motives or my reasons for action, my behavior largely will be regulated by a complex of forces at work *in* or *on* me, just as the behavior of a young child or a non-human animal is regulated by such forces. When, however, I take the backward step, I can self-consciously reflect[8] upon my desires and beliefs and decide whether such desires or beliefs are worth having for reasons I can endorse, and I can amend or even produce new beliefs or desires. Once I have thus either produced a new belief or desire, or amended or endorsed an existing belief or desire, I will have thereby taken charge of or assumed ownership of it, such that my subsequent actions will flow from and be attributable to *me*. Third, the capacity for self-composition or self-improvement. Before I have reflected upon myself, I will possess little understanding of what kind of person I am, what my personal attributes are, how they hold together, or what I am capable of. I may, for example, associate myself with a vague and fluctuating set of attributes: "I am fairly intelligent, lazy, compassionate, and selfish, etc." Yet, once I reflect on myself, not only can I achieve a better grasp of my nature, but, fortified with

[7] I take the "deep" qualification from Korsgaard (2009, p. 19-20) and will elaborate on it in Chapter 5.

[8] I specify "self-consciously" reflect because I do not rule out the possibility that one can reflect un-self-consciously—that is, that one can reflect on oneself without knowing one is even doing so.

this knowledge, I can take steps to improve myself. I can, for example, cultivate features I approve of or minimize those I dislike and I can better coordinate or integrate such elements and, in this way, gradually transform myself into the kind of person I wish to be.

The aspiration to achieve and exercise such self-oriented powers should not strike one as exotic or mysterious. Anyone who has sat on a psychologist's couch, who has asked himself "Why do I feel like this?" or "Why did I behave like that?"—who has tried to curb a habit or otherwise alter his behavior or character, has wished to exercise such powers. Indeed, the overwhelming appeal of the modern self-empowerment movement rests on the assumption that such powers are readily available to us.[9] Yet one may wonder to what extent such powers *actually* exist. And in virtue of what *exactly* does the assumption of reflective distance afford us access to them?

These questions will constitute the primary subject of this thesis. But here, in broad outline, is the Rationalist picture I will develop and, in large part, refute. Let us start with Kant, who tells us: "The fact that the human being can have the "I" in his representations [i.e., is self-conscious] raises him infinitely above all

[9] See Guignon: "Self-help and human potential movements reinforce the faith in control […] by pressuring individuals to take control of their own lives through self-inspection, self-surveillance, and self-assertion" (2004, p. 166). This sort of mentality stands out with exceptional vividness in the work of leading self-help guru, Tony Robins, who reports that "using the power of decision gives you the capacity to get past any excuse to change any and every part of your life *in an instant*. It can change your relationships, your working environment, your level of physical fitness, your income, and your emotional states. It can determine whether you're happy or sad, whether you're frustrated or excited, enslaved by circumstances, or expressing your freedom. It's the source of change within an individual, a family, a community, a society, our world. […] In a moment you can seize the same power that has shaped history." Anthony Robins, *Awakening the Giant Within* (Free Press, 2007), p. 35.

other living beings on earth. Because of this he is a *person* […] i.e., through rank and dignity an entirely different being from *things*, such as irrational animals…"[10] That is to say, for Kant, our capacity to represent ourselves—to regard ourselves as an "I," renders us not only superior to or "above" crude *things* (the set of which includes "irrational animals"), but *infinitely* above them, such that the difference must be understood not as a matter of *degree* but of *kind*. That is, by way of self-consciousness, a *new entity*—a "person"—springs into being, and persons, unlike all other things and irrational creatures, possess dignity and are worthy of respect. Indeed, according to the Rationalist view, the assumption of self-conscious personhood endows human beings with a set of capacities or powers that are unavailable to mere things and irrational animals. Thus, Korsgaard tells us:

> (1) A non-human animal acts on what I call 'instinct.' Her instincts are her principles, and they constitute her will. […] You are not so lucky. As a rational agent, you are aware of the grounds of your beliefs and actions—or, I should say, the potential grounds. For being aware of them gives you some distance from them and puts you in control. Self-consciousness divides you into two parts, or three, or any number of parts you like… [And] now you must pull yourself together by making a choice.[11]
>
> (2) Our capacity to turn our attention on to our own mental activities is also a capacity to distance ourselves from them. I perceive, and I find myself with a powerful impulse to believe. But I back up and bring that impulse into view and then I have a certain distance. *Now the impulse doesn't dominate me* and now I have a problem. Shall I believe? Is this perception really a reason to believe? I desire and I find myself with a powerful impulse to act. But I back up and bring that impulse into view and then I have a

[10] Immanuel Kant, *Anthropology from a Pragmatic Point of View*, Robert Louden (ed.), (Cambridge University Press, 2006) p. 15, §127.
[11] Christine Korsgaard, *Self-Constitution Agency, Identity, and Integrity*, (Oxford University Press, 2009), p. 212-213.

> certain distance. *Now the impulse doesn't dominate me* and now I have a problem."[12]

On Korsgaard's account, a non-human animal's behavior is unreflectively regulated by its instincts, is, as she puts it, "normatively loaded"[13]. But we human beings are "not so lucky." Rather, by means of my capacity for self-consciousness I can become aware of my instincts and "mental activities". Insofar as I have become conscious of my instincts and mental activities, I will have assumed a reflective "distance" with respect to them. In virtue of having achieved this distance, my instincts and mental activities cannot *compel* me to act or to believe. Rather, *I* will have assumed power over *them*, such that I can determine whether they are *worthy* of endorsing or following up on in virtue of my *reasons*.[14] As Nagel puts it, "One is suddenly in the position of judging what one ought to do, against the background of all one's desires and beliefs, in a way that does not merely flow from those desires and beliefs but operates *on* them" (1997, p. 110, emphasis mine). That is, having assumed a reflective stance, the outcome of my deliberations will ostensibly *not* flow from or be determined by "the background" of my existing desires and beliefs; rather, *I* will bring about or author my deliberative conclusion *in virtue of my own self-legislative authority*. Velleman, similarly, tells us:

> [T]he agent's role cannot be played by any mental states or events whose behavioral influence might come up for review in practical

---

[12] Christine Korsgaard, *The Sources of Normativity*, (Cambridge: Cambridge University Press, 1996), p. 93, italics mine. (Henceforth: *SN* or 1996b)

[13] Korsgaard, *The Constitution of Agency* (Oxford: Oxford University Press, 2008), p. 3-4.

[14] See also Korsgaard (1996, p. 100): "When you deliberate, it is as if there were something over and above all of your desires, something which is *you*, and which *chooses* which desire to act on" (emphasis hers).

> thought at any level. And the reason why it cannot be played by anything that might undergo the process of critical review is precisely that it must be played by whatever *directs* that process. The agent, in his capacity as agent, is that party who *is always behind, and never solely in front of the lens of critical reflection*, no matter where in the hierarchy of motives it turns.[15]

Like Korsgaard and Nagel, Velleman regards the role of the reflective agent as something that "cannot be played by" our mental states or events (at least those that one can subject to review); rather, the agent himself must *stand behind* and *direct* such mental states and events. Finally, consider Moran, who baldly proclaims that: "[A] non-empirical or transcendental relation to the self is ineliminable."[16]

The general idea, according to the Rationalist Story, might thus be summarized as follows. My empirical self may be defined and constrained by a set of conditions built into, or extending from, its nature. However, when I assume a self-conscious, reflective stance *with respect to* my empirical nature, a distance opens up between the subjective "I" and the empirical "me". This distance to some degree both *exempts* the "I" from the conditions that define and constrain the "me" and *affords* the "I" a unique set of epistemic, agential, and constitutional powers that it can exercise *with respect to* the "me". Indeed, according to Korsgaard, once I have fully deployed such powers, I can assume

---

[15] David Velleman, "What Happens When Someone Acts" from *The Possibility of Practical Reason* (Oxford University Press, 2000), p. 139, italics mine.

[16] Richard Moran, *Authority and Estrangement* (Princeton University Press, 2001), p. 90. See, again, Nagel: "The objective self functions independently enough to have a life of its own. It engages in various forms of detachment from and opposition to the rest of us, and is capable of autonomous development. […] [I]t places us both inside and outside he world, and offers us possibilities of transcendence which in turn create problems of reintegration." Thomas Nagel, *The View From Nowhere* (Oxford University Press, 1986), pp. 65-66.

complete possession and control of myself, such that I will be, "entirely self-governed, so that all of [my] actions, in every circumstance of [my] life, are really and fully [my] own: never merely the manifestations of forces at work in [me] or on [me]…. [I will have become] completely self possessed…"[17]

Again, it is easy to appreciate the seductive appeal of such a model of consummate, captain-like, self-possession. To begin with, such an account seems to track our own phenomenology: it really *feels like* we can "get a-hold of ourselves," "collect ourselves," and "direct ourselves." Moreover, the achievement of such a state would appear to be of enormous practical value—would allay one's fears of, for example, being mystified or surprised by one's own behavior, succumbing to unwelcome temptation, or being helplessly plagued by undesired personal desires or beliefs. Indeed, being the self-reflective, socially conscious, self-judging creatures we are, perhaps we *cannot help* but aspire to some version of Korsgaard's ideal of self-possession. Yet in the following chapters, I will argue that such a longed-for ideal of self-possession is not only metaphysically implausible and therefore untenable,[18] but that one's efforts to achieve it are often self-undermining and that the practical and ethical ramifications of such an ambition are, in many respects, unwholesome.

I will proceed as follows. In Chapter 2, invoking the work of James and Sartre, I will more fully flesh out the "Rationalist" or "Captain of Crew" model of

[17] Christine Korsgaard, "Self-Constitution in the Ethics of Plato and Kant," *The Journal of Ethics* (1999), 3: 1-29, p. 22. Note that Korsgaard does not employ the "I," "me" terminology, but the suggestion in the quoted passage is that the "I" of self-reflection will have taken full possession of all the attributes of the "me."

[18] Or, to invoke Blackburn's less amicable characterization—such an ideal is a "romantic, existentialist fancy." See Simon Blackburn, *Ruling* Passions, (Oxford University Press, 1998), p. 257 and p. 260.

the self, and, in doing so, spell out the ostensive roles played by the "I" and the "me" with respect to the collective or corporate "self." In presenting this model, I will introduce and characterize the putative *unilateral* influence that the reflective "I" can exercise with respect to the facts of the empirical "me". I will conclude this chapter with an analysis of the various self-referential terms to be employed over the course of the thesis, and briefly review some of the interpretive difficulties to which these terms inevitably give rise. In Chapters 3-5, I will challenge Part (4) of the Rationalist Story; that is, I will challenge the notion that reflective distance affords us the aforementioned set of epistemic, agential, and self-compositional powers. Note, importantly, that while I will *defend* the thesis that our capacity for self-consciousness *does* afford the reflective "I," ***as provisionally understood***, such powers, the *nature* and *extent* of such powers will in every case be constrained and conditioned by the facts of the "me". As such, I will argue that, when correctly understood, the "I" *never* acts "on," or "over" the "me." Rather, the operations of the "I" should be understood as *expressions* or *products* of the facts of the "me". This will lead in Chapter 6 to an all-out repudiation of the Rationalist, Captain-of-Crew model of the self and a discussion of what I regard as a far-more-plausible, Humean, "Holistic," or, what I shall call, "Crew-of-Captains" model, where the "I" and the "me" should be understood *not* as two distinct entities, but rather as two aspects of one "corporate" entity. Finally, in Chapters 7 and 8, I will address some of the metaphysical and ethical merits of the Holistic model of the self. That is, I will suggest that the Holistic model offers an account of the self and of human behavior that makes much more

metaphysical sense and I will suggest that the Holistic model also allows for a more accommodative, compassionate and wholesome view of human nature.

CHAPTER 2

THE RATIONALIST MODEL OF THE SELF: ME, MYSELF AND I

2.1 James and Sartre

A helpful entry point to the Rationalist, or the "Captain-of-Crew" model of the self will be William James' characterization of what he elsewhere calls the "Duplex Self."[19]

> We may sum up by saying that personality implies the incessant presence of two elements, an objective person, known by a passing subjective Thought and recognized as continuing in time. *Hereafter let us use the words ME and I for the empirical person and the judging Thought.* (1918, p. 371, emphasis his)

James' self can thus be understood in terms of these two elements, the "I" and the "me" (I shall express the latter in lower-case). The "I" he associates with the "thinker," the "subject" of thought, or "pure ego," which, he tells us "at every moment *is* conscious" (2001, p. 62, italics his). The "me," or "empirical ego," on the other hand, is the *object* of thought, that is, the part of the self of which the "I" is, or can be, conscious. As James puts it: "*In its widest possible sense […] a man's Me is the sum total of all that he* CAN *call his, not only his body and his psychic powers, but his clothes and his house, his wife and children, his ancestors and friends, his reputation and works, his lands and horses and yacht and bank account*" (2001, p. 44, emphasis his). While the "I" is the thinker and the observer of the "me," it is also involved in "attending to" and "appropriating"

[19] In Chapter 6, I will trace Rationalism to much earlier roots.

aspects of the "me"—remembering certain elements, caring for them and letting go of those elements for which it no longer cares. (2001, p. 82)

Let us now turn to Sartre's account of the self, which in many respects resembles James' construal but will more closely approximate the full-blown Rationalist or "Captain of Crew" model that will be the target of my critique.[20] Like James, Sartre divides the self into third and first-personal, aspects. On the third-personal front, Sartre offers the "factical" aspect, what he calls the "in-itself". This factical aspect, like James' "me," is composed of one's facts—characteristics, dispositions, past actions—items whose being, as he puts it, "is full of itself."[21] On the first-personal front, Sartre offers his version of the "I," what he calls the "for-itself," that "transcends" the facts and, as such, can never be identical to or "coincide with" such facts[22]. Further, like James' "I," Sartre's transcendental element can attend to and appropriate the facts of "me". But Sartre's "I" can exercise a more robust agency with respect to the "me" than

[20] The following review does not pretend to do justice to the complexity and nuance of Sartre's treatment. It is only intended to present, in broad outline, Sartre's two central aspects of the self. Note also that Sartre does not strictly employ the terms "I" and "me," as James does, to refer to the different aspects of the self, but I believe these terms can be innocently imported.

[21] "The in-itself is full of itself, and no more total plenitude can be imagined, no more perfect equivalence of content to container. There is not the slightest emptiness in being, not the tiniest crack through which nothingness might slip in." (Jean-Paul Sartre, *Being and Nothingness*, Jean-Paul Sartre, Hazel E. Barnes (trans. and intro.), (Washington Square Press: New York, 1956), p. 120-121.

[22] See, for example, Sartre's treatment of the paederast in his discussion of "bad faith." "He would be right actually if he understood the phrase 'I am not a paederast' in the sense of 'I am not what I am.' That is, if he declared to himself, 'To the extent that a pattern of conduct is defined as the conduct of a paederast and to the extent that I have adopted this conduct, I am a paederast. But to the extent that human reality can not be finally defined by patterns of conduct I am not one'" (Sartre, 1956, p. 108). Here, "patterns" of conduct may be understood as elements of one's facticity.

James allows[23]. That is, Sartre's "I" can *establish* the correct interpretation of the facts of the "me," it can *alter* the present facts of the "me" and it can *determine* the future facts of the "me". As Sartre puts it: "It is I, always I, according to the ends by which I illuminate these past events. Thus all my past is there pressing, urgent, imperious, but its meaning and the orders which it gives me I choose by the very project of my end" (1956, p. 640). Indeed, Sartre's for-itself can *unilaterally* influence the facts of the "me". That is to say, Sartre's "I" can influence the "me" in ways that are not conditioned or governed by the facts of the "me."

2.2 On The Nature of Unilateral Influence

Indeed, I regard such putative unilateral influence—that is, the possibility that the "I" can influence the "me" in ways that are not attributable the facts of the "me"— as the *cornerstone* of the Rationalist position. For this reason, I will need to spend a little time clarifying the idea, both its nature and its particular relevance to the Rationalist position.

Recall, first, how Velleman characterizes the agent as "that party who *is always behind, and never solely in front of the lens of critical reflection*, no matter where in the hierarchy of motives it turns." Or recall Nagel's assertion that the agent's deliberation can proceed in a way that "does not merely flow from [one's] desires and beliefs but operates on them." Or, again, as Sartre has said: "It is I,

[23] James asserts that the "I" can "fix its attention" on a particular thought and in this way influence the course and therefore the outcome of thought, but he does not appear to attribute to the "I" the kind of productive or interpretive agency Sartre endorses. See James (2001), p. 82.

always I, according to the ends by which I illuminate these past events. Thus all my past is there pressing, urgent, imperious, but its meaning and the orders which it gives me I choose by the very project of my end". Or, finally, as Moran puts it: "[A] non-empirical or transcendental relation to the self is ineliminable." As such, on the Rationalist account, the behavior of the "I," at least in some respects, should *not* be understood as a *product* or *expression* of the facts of the "me"—not as something that "flows" or "derives" from the "me". Rather, the "I," at least in some respects, should be regarded as a *director* or *author* of such facts, the monarchic agent who autonomously *operates on or over* such facts.

But what might it mean, precisely, for the "I" to exert such unilateral influence over the "me"? That is to say, what conditions must be satisfied such that behavior of the "I" with respect to the "me" can be characterized accurately as "unilateral"? To properly understand the question, let us put it more generally, viz: What conditions must be satisfied such that it can be truly said of *any* x that it exerts unilateral influence over *any* y? Here it may be helpful to make us of a familiar illustration. Take two billiard balls, x and y. And imagine billiard ball x rolling across the table and striking ball y. We can easily observe[24] the causal influence that x will exert on y; that is, on force of impact, y will rebound away from x. More than this, the causal force that x brings to bear on y will not *derive from* y. That is to say, the force of x—understood as the product of its acceleration and mass—exists independently of y.[25] To better see this, note that

[24] Humean reservations notwithstanding.

[25] Granted, the extent to which and the manner in which y *responds* to the impact of x will depend on various features of y including its own state of motion and

we could remove y from the picture and substitute ball z, without in any way altering the set of causal powers x brings to bear. I suggest, thus, to the extent that the causal powers of x do not derive from y, that x will exercise a *unilateral* causal influence upon y.

We may therefore characterize such unilateral causal influence as follows:

> *Unilateral Causal Influence*: For any *x* and *y*, *x* exercises unilateral causal influence upon *y* insofar as the causal powers *x* bring to bear on *y* do not derive from *y*.

We are now better positioned to understand what I believe the Rationalist will need to demonstrate in order to successfully argue that the "I" can exert unilateral influence upon "me." That is, the Rationalist will need to show that the "I" can in fact influence the "me" in some way that is not altogether attributable to the facts of the "me". For, if the behavior of the "I" *were* wholly attributable to, or conditioned by, the facts of the "me,"—if, as Nagel expressed it, the behavior of the "I" entirely *flowed from* the "me"—then the behavior of the "I" would just constitute a further *expression* or *product of* such forces, not a *response to* or *reaction against* such forces.

Such is the Rationalist's challenge, as I see it. And in the proceeding chapters I will explore a number of accounts that purport to demonstrate such unilateral influence. All of these Rationalist accounts will, I admit, *powerfully* capture our phenomenology and will for this reason boast great intuitive appeal. In spite of this, I will argue that, in every instance, the powers of the "I" flow entirely from and are entirely conditioned by the "me". As such, I will argue that (1) the "I"

---

material constitution, but, again, the set of causal powers x brings to bear on y will not *derive* from y.

does not exercise unilateral powers with respect to the "me," and, further, (2) the "I" and the "me" should not be regarded as ontologically distinct.

2.3 Initial Clarification of Self-Referring Terms

Before proceeding, it will be necessary to shed some clarifying light on the set of self-referential terms that will be employed over the course of this dissertation. Note: I underscore that this will only constitute an *initial* clarification—and I cannot emphasize this strongly enough—for, as we proceed, the meanings of these terms will undergo continuous exploration and refinement, and, even after the most thorough analysis, the precise definitive boundaries of these terms will—of necessity, I shall argue—remain obscure and porous. Indeed, a central aim of this thesis will consist in showing why the definitions of these terms cannot be satisfactorily demarcated. Nonetheless, I hope that the following analysis will help alleviate some confusion.

Person and *"Person"*: I will use the term *person* or *persons* (no quotation marks) primarily to designate human beings, that is, to designate concrete human *people*. For example, I may point into a petting zoo populated by a number of pigs, donkeys, sheep and human visitors and ask, "How many persons are in the zoo?" If one human being occupied the zoo, the appropriate answer would be: "There is one person in the zoo." Now, I will also employ the term *"person"* (with quotations) in the Kantian, Rationalist sense to pick out the *subject* that comes into being via self-consciousness, that is, the self-regarding, self-reflecting,

agential "I".[26] As such, to return to our petting zoo, let us assume that one human occupant were to suddenly die. Under such conditions, assuming, as I will, that a functioning brain is a precondition for self-consciousness, the petting zoo would still contain one person, but it would not contain a "person".

Self: I will understand the word *self* to refer to the Jamesian or Sartrean *complex* of the self-reflecting "I" and the personal, factical features that the "I" has observed or can observe, the set of which comprises the "me". Let us call this complex a "psychological field." Thus, if you were to ask me to "describe my self," I would offer a list of all the attributes of this field I can consciously recognize, one of which is the fact that "I" observe them. Note, importantly, that the existence of my self is thus contingent on the existence of the self-reflecting "I" which observes this self. Thus, insofar as the existence of a reflecting "I" is contingent upon a living brain, a brain-dead person would *lack* a self.[27]

---

[26] Again, we have seen this account put forward by Kant, but see also Locke who defines a person as "a thinking intelligent Being, that has reason and reflection, and can consider it self as it self the same thinking thing in different times and places." (*An Essay Concerning Human Understanding*, (ed.) P. Nidditch (Oxford: Clarendon Press, 1979, p. 335) And Korsgaard: "The identity of a person, of an agent, is not the same as the identity of the human animal on whom the person normally supervenes" (2009, p. 19, and see also Korsgaard, 1996b, p. 102). See also Frankfurt: "To be a person, as distinct from simply a human organism, requires a complex volitional structure involving reflective self-evaluation." "The Faintest Passion," *Necessity, Volition and Love* (Cambridge University Press, 1999) p. 138. And see a similar account in Lynne Rudder Baker, *Naturalism And The First-Person Perspective* (Oxford University Press, 2013), pp. 147-156.

[27] So as not to bog down here, I will abstain from an exploration of the differences between "selves" and "persons". Though distinctions could be drawn, they would probably serve to confuse rather than clarify, and not much will turn on this distinction in my discussion. For an interesting treatment of such a distinction, see Velleman: "Identification and Identity" from *Contours of Agency*, *Essays on Themes from Harry Frankfurt*. (eds.) Sarah Buss and Lee Overton, (MIT Press,

Me and "me": The term *me* (without quotations) shall be used in the traditional sense, to pick out either the passive subject of my experience, e.g., "things happen to me," or my concrete human personhood: "the person in the picture is me". As discussed above, I will also employ the technical, quote-enclosed term *"me"* to refer to the entire set of my facts of which "I" can become conscious—my physical size, my past actions, my attitudes, relationships, etc.

Proper Names and *"Proper Names"*: Note that proper names, e.g., Jon (no quotes) will be used, generally, to refer to objective human persons, e.g., "The lost wallet belongs to Jon," or "The person in the photo is Jon." Note, however, that when it becomes important to specify a self-reflective "person," I will enclose the proper name in quotations. So, we might encounter the sentence: "Forced to decide which career to pursue, "Jon" asked himself which made the most sense." Or, "Jon was absentmindedly walking down the street when it suddenly struck him that he might have hurt his wife's feelings at breakfast. Why, "Jon" asked himself, did he have to be such a jerk?"

I *and "I"*: Finally, the terms I (no quotes), and the term "I" (quotes), will perhaps pose the greatest definitional challenge, so I emphasize the chiefly *prefatory* status of this analysis.

2002), p. 111. See also Velleman's discussion in "Self-to-Self," *Self to Self, Selected Essays*, (Cambridge, 2006). And see Catriona Mackenzie, "Bare Personhood? Velleman on Selfhood," *Philosophical Explorations*, Vol. 10, No. 3, September 2007.

Thus far I have primarily discussed the Jamesian "I" as that which signifies the subjective, agential locus of self-consciousness, that which observes and otherwise attends to the facts of the "me". But we also employ the term I (no quotations) to identify our objective human personhood; that is, we use the term to pick out the objective factive human being that we know ourselves to be.[28] To better see this, notice that someone may present a photograph of the petting zoo and point to the human person with thinning red hair and ask, "Who is that person in the petting zoo?" Now, I will examine the photograph and respond: "I am that person," where the term *I* is meant to pick out Jon Rosen, the objective human person standing among the pigs and the geese. Yet notice that I could have just as easily responded to the question as follows: "The person in the photograph is *me*—where, again, the terms *me* or "me" will also pick out the observable factual human creature, Jon Rosen. As such, in this case, the terms I, and me/"me", will refer to the very same objective creature. Yet notice, *critically*, that my very recognition of such facts that ["I am that person"] and ["That person is me/"me,""] must involve an act of self-conscious self-recognition. For, I, Jon Rosen, the human being, cannot recognize that the human being in the photograph is me/"me" unless "I" recognize that the image in the

[28] See Strawson: "[W]hen I think and talk about myself, my reference sometimes extends only to the self that I am, and sometimes it extends further out, to the human being that I am." Galen Strawson, *Selves: An Essay in Revisionary Metaphysics*, (Oxford: Clarendon Press, 2009), p. 31. See also Liu's articulation of this problem: "In self awareness or self-knowledge, both uses [of the I] seem to be present. "*I* believe that *I* am the tallest person in the class"; "*I* know that *I* am not sad about her departure." How can there be two selves indicated in these self-reports, or is it just one self who knows, perceives, thinks about or is aware of, *the same self*? How can the self be both the knower and the known?" JeeLoo Liu and John Perry (eds.) *Consciousness and the Self, New Essays* (Cambridge, 2012), p. 3.

photograph corresponds to a set of facts with which "I" self-identify. As such, my active employment of both pronouns *I* and *me/"me"* to designate my objective identity will *reflexively implicate the subjective, self-identifying "I"*. Or, to put the matter a bit differently, my recognition of my objective existence *as me* necessarily implicates my recognition of my subjective-reflective existence as *"I"*.[29]

As such, our terminological boundaries have already begun to blur. Yet this, as I suggested above, and will go on to argue, is *exactly what is to be expected*. Indeed, in the proceeding chapters I will explore a series of accounts where, on the Rationalist view, the self-reflective "I," operating as the agential subject of self-consciousness, is understood to unilaterally influence the facts of the "me". In each case, I will argue that the Rationalist mischaracterizes what is actually going on. That is, in each case I will argue that the behavior of the "I" is *entirely conditioned* by the facts of the "me," and, as such, that we cannot truthfully speak of the "I" unilaterally influencing the facts of the "me". Indeed, as I shall eventually—if perhaps not as successfully—argue, the "I" and the "me" should not be regarded as distinct entities, at all.

[29] See John Perry's wonderful elaboration on this tricky and fascinating point in "The Problem of the Essential Indexical," *Nous*, Vol. 13, No. 1 (March 1979), pp. 3-21.

CHAPTER 3

SELF-KNOWLEDGE

3.1 Introduction

"I must first know myself," Socrates tells us, "as the Delphian inscription says; to be curious about that which is not my concern, while I am still in ignorance of my own self, would be ridiculous."[30] Socrates maintained that self-knowledge was of supreme importance because without it one could not cultivate a good soul or lead a virtuous life. Perhaps our contemporary interest in self-knowledge is not as virtue-oriented. As Guignon puts it, we seek self-knowledge today not as much in the service of goodness or honor, but in order to achieve the dignity of becoming bounded, masterful, and autonomous. In either case, we regard self-knowledge as a requirement on becoming the type of person we wish to be and for living the kind of life we wish to live.

Luckily, we are not hopelessly in the dark about ourselves.[31] Indeed, some philosophers have suggested that when it comes to self-knowledge—as compared,

---

[30] Plato, *Phaedrus*, from *Selected Dialogues of Plato*, trans. Benjamin Jowett (Modern Library Classics, 2001), 230 a-b.

[31] Contrary to views expressed by, for example, Nietzsche: "And so we necessarily remain a mystery to ourselves, we fail to understand ourselves, we are bound to mistake ourselves. Our eternal sentence reads: 'Everyone is furthest from himself'—of ourselves, we have no knowledge." Nietzsche, *Genealogy of Morals*, (Oxford University Press, 2009) p. 5. And Camus: "For if I try to seize this self of which I feel sure, if I try to define and to summarize it, it is nothing but water slipping through my fingers. I can sketch one by one all the aspects it is able to assume, all those likewise that have been attributed to it, this upbringing, this origin, this ardor or these silences, this nobility or this vileness. But aspects cannot be added up. This very heart which is mine will forever remain indefinable to me." Albert Camus, *The Myth of Sisyphus and Other Essays* (trans. and ed.) Justin O'Brien (New York: Vintage, 1983), p. 19.

for example, to knowledge of the outside world—we enjoy infallibility. Consider Descartes, for example, who "affirm[s] with certainty that there can be nothing within [him] of which [he is] not in some way aware."[32] Or Rousseau, who declares:

> I have but one faithful guide on which I can depend: this is the chain of sentiments by which the succession of my existence has been marked, and by these the events which have been either the cause or the effect of the manner of it. […]. I may omit facts, transpose events, and fall into some errors of dates; but I cannot be deceived in what I have felt, nor in that which from sentiment I have done that have marked the development of my being.[33]

And Mill tells us that, "with respect to his own feelings and circumstances, the most ordinary man or woman has means of knowledge immeasurably surpassing those that can be possessed by anyone else."[34]

Of course such rosy assessments of introspective transparency have fallen out of popular favor. Freud's investigations into the unconscious cast a sobering light on our prospects for self-knowledge[35] and a surfeit of empirical studies testifies to

---

[32] Rene Descartes, *First Reply to Objections*, from *The Philosophical Writings of Descartes Vol. II*, trans. John Cottingham, Robert Stoothoff, Dugald Murdoch, (Cambridge University Press, 1985) p. 77 (107). Though, Descartes appears to have had some reservations: "[M]any people do not know what they believe, since believing something and knowing that one believes it are different acts of thinking, and the one often occurs without the other." *Discourse on the Method* from *The Philosophical Writings of Descartes Vol. II*, (trans.) John Cottingham, Robert Stoothoff, Dugald Murdoch, (Cambridge U. Press, 1985), p. 122 (23).

[33] Jean Jacques Rousseau, *The Confession of Jean Jacques Rousseau* Vol. 1 (London: Privately Printed for Members of the Aldus Society), Book VII, found in Guignon, (2004) p. 68.

[34] John Stuart Mill, *On Liberty* (London: W. Parker and Son, West Strand, 1859), Section (IV.4), p. 137.

[35] See, for example, Freud: "The unconscious is the larger circle which includes within itself the smaller circle of the conscious. […] [T]the unconscious is the real psychic; *its inner nature is just as unknown to us as the reality of the external world, and it is just as imperfectly reported to us through the data of the consciousness as is the external world through the indications of our sensory*

the fact that our access to the workings of our own minds is neither as immediate nor as trustworthy as we might like or expect. Nonetheless, we generally sustain a sanguine optimism with respect to our capacity for self-understanding. And, indeed, such optimism is at least partly well founded. For, insofar as we are self-conscious, we at least possess *access* to a kind of knowledge that, by definition, un-self-conscious creatures are denied. Nonetheless, in this chapter, I will make the case that the alleged epistemic powers and privileges of the reflective "I" are far less robust than our optimism warrants. Indeed, I will argue that the "I" (1) Overestimates the accuracy with which it can identify the existing facts of "me"—read such feelings, desires, or beliefs, so to speak, "off the page"; (2) Overestimates the extent to which it can exercise "rational authority" in the service of determining or "authoring" desires and beliefs; and (3) Overestimates the extent to which the "I" can *render true* certain features of the "me" through the act of "endorsing" or "identifying" with them. In each case, I will argue that the epistemic powers that the "I" brings to bear on self-analysis are pervasively influenced by the facts of the "me."

3.2 On Reading Off The Page: The Accuracy of Self-Directed Attention

We are fortunate to be endowed with perceptual mechanisms that deliver reliably accurate readings of the external world. If this were not the case, we would be incapable of successfully navigating the external world: we would bump into objects we were not seeing correctly, would consistently fail to pick up on

organs." Sigmund Freud, *Dream Psychology: Psychoanalysis For Beginners*, (trans.) M.D. Eder, (intro.) Andrew Tridon, (New York: The James A. McCann Company, 1920), p. 224, emphasis his.

auditory cues, etc.[36] Of course it sometimes happens that our impressions of the external world are inaccurate or misleading. Suppose, on a brilliantly sunny day, I see what resembles an odd-looking man standing at the side of the road. I focus more attention on the object and I realize that this "man" is really just two tall mailbox posts standing side-by-side. On such occasions, I am doubly grateful for the reliable functioning of my perceptual mechanisms: not only are they generally reliable, but, with a focusing of attention, these mechanisms reliably allow me to overcome false impressions.[37]

Can we say the same for self-perception[38]? In many respects, it seems our self-perceptions are also reliably accurate. This must be the case, because otherwise we would be strangers to ourselves: we would not be able to know when asked, for example, whether we were thirsty or tired, happy or sad, or whether we preferred country music over heavy-metal. As Timothy Wilson puts it: "It would be extremely maladaptive to be confused about whether or not we feel pain when

[36] Note that I'm making a *very modest* claim here. I am not suggesting that our perceptual mechanisms *always* deliver reliable information, or that they render us capable of understanding the "true nature" of the world, or even that such mechanisms provide reason to believe that the external world is not a mirage or a dream. I am simply making the uncontroversial claim that our ability to successfully navigate the world (whatever it is), to pick up on sensory cues, to identify objects, and so forth, is ensured by our reliably-functioning mechanisms of perception.

[37] I emphasize in most, but not all, cases. See, for example, the Muller-Lyer illusion and others cases, where regardless of my increase in attention, or alteration in perspective, I cannot overcome my misperception. A number of such visual illusions are provided here: http://www.michaelbach.de/ot/.

[38] In using the word "perception" here, I am not expressing any preference for a particular theory of self-knowledge. I am employing the term only to express the uncontroversial proposition that we are capable of achieving conscious access to, i.e., "perceiving," our own desires, beliefs and sensations.

touching a hot stove or fear when confronted by a mugger in a dark alley."[39] Of course, just as we can misperceive aspects of the external world, we can also misperceive aspects of ourselves. We may, for example, misread a feeling we are experiencing—perhaps envy for the achievement of a friend, or an inappropriate romantic attraction. And yet, when we are in doubt about such things, we believe we can "step back" and take a "good hard look" at ourselves and, in doing so, reliably overcome such misperceptions, just as we can overcome misperceptions of the external world.

The common-sense intuition that the mechanics of self-knowledge mirrors the mechanics of object knowledge has been nicely expressed in what Silvia and Gendolla[40] call the *Perceptual Accuracy Hypothesis* (PAH)—a hypothesis with which Silvia and Gendolla *disagree* but which they acknowledge as "the least controversial issue in contemporary self-awareness research."[41] PAH makes two principle claims. Claim (1): The more attention we give to an external object, the better we come to see, and therefore to know, that object. Silvia and Gendolla elaborate:

---

[39] From "Strangers to Ourselves: The Origins and Accuracy of Beliefs about One's Own Mental States," from *Attribution: Basic Issues and Applications*, (ed.) John H Harvey and Gifford Weary (Academic Press: 1985).

[40] See Paul J. Silvia and Guido H. E. Gendolla, "On Introspection and Self-Perception: Does Self-Focused Attention Enable Accurate Self-Knowledge?" *Review of General Psychology*, (2001), Vol. 5, No. 3, pp. 241-269.

[41] Silvia and Gendolla, (2001), p. 242. Silvia and Gendolla cite T.S. Duvall and R.A. Wicklund as early proponents of this view. See T.S. Duval and R. A. Wicklund, *A Theory of Objective Self-Awareness* (New York: Academic Press, 1972). They trace the first explicit formulation of the view to, e.g., F.X. Gibbons, C.S. Carver, M.F. Scheier, and S.E. Hormuth, "Self-focused attention and the placebo effect: Fooling some of the people some of the time," *Journal of Experimental Social Psychology*, (1979), 15, pp. 263-274.

> [A]ttention allows selective and detailed information about stimuli. When attention is guided to an object, the object becomes figural against the background of the perceptual field. Knowledge of the object thus becomes more detailed and clarified, and the perception of the object is consequently more accurate.[42]

Consider the mailbox example mentioned earlier. On first glace, I am pretty sure I see a man standing on the side of the road. In order to verify the accuracy of my perception, I take a closer look: I focus more attention on this odd-looking man. The focusing of my attention further engages my perceptual mechanisms, thus permitting them to pick up on more informative[43] details, and causing the less informative features to diminish or "recede" into the background. The second claim made by the Perceptual Accuracy Hypothesis is, (2): Object perception and self-perception are "dynamically identical." That is to say, just as the focusing of one's attention on an external object renders one's perceptual mechanisms better able to make out informative features of that object, thus the focusing of our introspective attention on our personal characteristics generally ensures that our self-scanning mechanisms arrive at more accurate readings.

Yet as Silvia and Gendolla and other critics of PAH have pointed out, research on introspection simply does not bear out this thesis. Indeed, critics of PAH cite evidence demonstrating that, in a very wide range of cases, self-directed attention not only *fails* to enable us to accurately identify our existing attitudes, but that it

[42] Silvia and Gendolla, (2001), p. 242. Silvia and Gendolla speak specifically of visual attention, but I see no reason why PAH wouldn't extend to auditory or olfactory attention, or any attention that has been focused for the purpose of achieving a clear perception. For example, when I listen more closely to a poorly recorded conversation, I can better make out what the voices are saying—the static recedes into the "background" of my attention, and the distinct words loom to the "surface".

[43] By "informative," I mean, truth bearing, or veridical—that is to say, details that will enable me to identify the actual nature of the object in question.

can serve to reinforce false impressions, and even cause us to override more accurate, spontaneous or unreflective assessments.[44] Again, the evidence is quite extensive, so I'll just touch on some of the highlights. Consider, first, cases of *affective ignorance*, that is, instances where we misread our own presently occurring emotional states. True, we may not experience difficulty in identifying instances of shock or horror. But we often fail to identify feelings such as envy, anger, or arousal[45] and we very commonly fail to identify more "background" feelings of anxiety and depression which, as Haybron points out, are "comparatively diaphanous, offering us not so much distinct objects within the field of consciousness as alterations of the field itself, coloring the entirety of our experience".[46] Consider also the extent to which people fail to acknowledge their own racial prejudice. Implicit bias tests, for example, routinely disclose the hidden or "automatic" biases that imperceptibly influence the judgments of self-

---

[44] For a nice review of such studies, see Shelley E. Taylor and Jonathan D. Brown, "Illusion and Well-Being: A social Psychological Perspective on Mental Health." *Psychological Bulletin* (1988), Vol. 103, No. 2, pp. 193-210. See also, Richard E. Nisbett and Timothy DeCamp Wilson, "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review*, (1977), May, Vol. 84, No. 3; For an interesting discussion of our phenomenology of sight and thought, see Eric Schwitzgebel, "The Unreliability of Naïve Introspection", *Philosophical Review*, (2008) Vol. 117, No. 2. See also, Alison Gopnik "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality," from Alvin Goldman, ed., *Readings In Philosophy & Cognitive Science*, (MIT Press, 1993), pp. 315-346.

[45] See Eric Schwitzgebel (2008). See also a study done on the relationship between homophobia and homosexual arousal: H.E., Adams, L.W.J. Wright, and B.A. Lohr, "Is Homophobia Associated With Homosexual Arousal?" *Journal of Abnormal Psychology*, 105, 1996, 440-445.

[46] Daniel M. Haybron, "Do We Know How Happy We Are? On Some Limits of Affective Introspection and Recall," *NOUS* 41:3 (2007), p. 398. Haybron cites recent philosophical studies of mood including C. Armon-Jones, *Varieties of Affect*, (Toronto: University of Toronto Press, 1991), and P.E. Griffiths, *What Emotions Really Are*, (Chicago: University of Chicago Press, 1997).

proclaimed non-racist white subjects.[47] Consider, further, studies that explore our "misattribution" of feelings and preference, that is, both our failure to recognize and our tendency to confabulate underlying causes of our feelings and preferences. In one study, subjects were asked to rank four identical pairs of stockings set on a shelf. Overwhelmingly, subjects displayed a preference for the stockings set further to the right. When subjects were asked if the positioning of the stockings influenced their preference, they vigorously denied it.[48] Subjects have also demonstrated a pervasive failure at recognizing changes in their attitudes, an inability to recognize how they arrive at solutions to creative problems, and an inability to recognize their propensity for "subjective optimization" that is, the tendency to retrospectively re-construe or "paint" disappointing circumstances as preferable or ideal.[49]

Of course, just as one can usually take a harder look at an object and thereby overcome a misperception, one might therefore expect that the subjects of such studies, when informed of the common susceptibility to such biases or causal-misattributions, could take a *harder* look at their own features and detect (and thereby overcome) their errors. Yet, studies have shown that, even when so informed, subjects not only routinely fail to detect their biases, but they

[47] A variety of such tests are available on-line at Harvard's "Project Implicit": https://implicit.harvard.edu/implicit/aboutus.html. Wilson offers a nice summary of these studies (2002), p. 188-194.

[48] See Richard E. Nisbett and Timothy DeCamp Wilson, "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review*, Vol. 84, No. 3 (1977), p. 243. See also the "Woman on the Bridge" experiment as described in Donald G Duttan and Arthur P. Aron, "Some Evidence for Heightened Sexual attraction Under conditions of High Anxiety." *Journal of Personality and Social Psychology* 30, no. 4 (1974): 510-17.

[49] For a nice overview of the empirical data supporting these findings, see Nisbett and Wilson (1977).

strenuously continue to *deny* that their judgments have in any way been thus influenced or undermined.[50] That is to say, due to their "bias blind spots," such people's biases often render them incapable of recognizing their own biases. As Blackburn has put it: "We can compare the situation to looking at our own eyes in a mirror. We might see that our eyes are cloudy; but if they are, it will be with cloudy eyes that we see it." (1998, p. 261)

3.2.1 On Distance and Objectivity

Why, as such studies indicate, does increased introspective attention apparently fail to yield the epistemic advantage we expect it should? For insight into the question, let us return to the Rationalist Story, and consider some foundational assumptions that I suggest underlie our confidence in the proposition that self-directed attention *should* afford such an advantage. First, consider Velleman's assertion that

> consciousness just seems to open a gulf between subject and object, even when its object is the subject himself. Consciousness seems to have the structure of vision, requiring its object to stand across from the viewer—to occupy the position of *Gegenstand* [object or thing] (2008, p. 179).[51]

In order to see something with our physical eye, Velleman explains, the object of sight must stand "across from" the eye. Likewise, in order for something to become an object of our "inner-vision" this object will need to stand *before* the

[50] See Emily Pronin, Daniel Y. Lin and Lee Ross, "The Bias Blind Spot: Perceptions of Bias in Self Versus Others," *Personality and Social Psychology Bulletin*, (2002) March, Vol. 28, No. 3, pp. 369-381.
[51] J. David Velleman, "The Way of the Wanton," from *Practical Identity and Narrative Agency*, (eds.) Kim Watkins and Catriona, (Routledge, 2008), p. 179.

introspective eye.[52] Now notice how Korsgaard builds on these themes of distance and objectivity. Recall the passage quoted above:

> Our capacity to turn our attention on to our own mental activities is also a capacity to distance ourselves from them. I perceive, and I find myself with a powerful impulse to believe. But I back up and bring that impulse into view and then I have a certain distance. *Now the impulse doesn't dominate me* and now I have a problem. Shall I believe?

Like Velleman, Korsgaard asserts that our very capacity to attend to our mental activities requires that a "distance" open up between the viewer and her mental activities. Indeed, on Korsgaard's account, this distance not only allows us to bring our mental activities into view, but it allows us to view them with a certain *detachment*, such that they no longer "dominate" us. The suggestion is that, absent such distance, nothing would stand in the way of our impulses dictating our beliefs or actions, just as they dictate an irrational animal's beliefs and behaviors[53]. Finally, consider how Nagel further amplifies this notion of distance and objectivity. He asks:

> How do I abstract the objective self from the person TN? By treating the individual experiences of that person [himself, TN] as data for the construction of an objective picture. I throw TN into the world as a *thing* that interacts with the rest of it, and ask what the world must be like from no point of view in order to appear to him as it does from his point of view. For this purpose my special link with TN is irrelevant. Though I receive the information of his point of view directly, I try to deal with it for the purpose of constructing an objective picture just as I would if the information were coming to me indirectly. […] [I]n a general way, I try to do with his perspective on the world what I could do if information

[52] Here I take Velleman to be making a conceptual claim: that is, no *x* can perceive a *y* if no distance stands between an *x* or a *y*, for if no distance stood between an *x* and a *y*, *x* could not achieve a perspective from which it could perceive *y*.

[53] Recall, as Korsgaard puts it, an animal's impulses or instincts are "normatively loaded" (2008, p. 3-4).

> about it were reaching me from *thousands of miles away* not pumped directly into my sensorium but known from *outside*.[54]

In order to see himself objectively, as a "thing," Nagel "throws himself into the world". Once he has done so, Nagel's "special link" with TN will have become "irrelevant," such that TN's experiences will not be "pumped directly into [TN's] sensorium," but will come to him in the form of impersonal "data" from "thousands of miles away," or, as he puts it from the "outside". Indeed, from such an "outside" perspective, as Silvia and Gendolla have put it, "[t]he self is viewed from the perspective of an actual, hypothetical, or generalized *other*. The self is seen as an object in the world distinct from others, an object with boundaries, fixed properties and the capacity to be controlled" (2001, p. 243, emphasis mine). Thus, just as increased attention reliably allows us to better make out the defining boundaries and detect the veridical details of external objects—thus do we expect that our ability to view ourselves from a distance, to "throw ourselves into the world," should allow us to achieve a more accurate and objective perspective on ourselves. Yet, given the abovementioned empirical evidence, it is clear that the taking of a backward step *does not* yield more accurate introspective knowledge anywhere near as commonly as object knowledge. What has gone wrong? For insight into the matter, it will be helpful to review some common sources of distortion, both non-motivational and motivational.

[54] Thomas Nagel, *The View From Nowhere* (Oxford University Press, 1986), p. 62, emphasis mine.

3.2.2 Sources of Distortion – Non-Motivational and Motivational

Consider, first, non-motivational aspects of our psychology. Recall again Haybron's characterization of the influence of our moods, suggesting that they alter *the entire* field of our consciousness, and, as such, fail to constitute "distinct objects" that can be located within the field (2007, p. 398). A background, unrecognized feeling of malaise, for example, can "sour one's experience far more than the sharper and more pronounced ache that persists after having stubbed one's toe" (2007, p. 398). Indeed, studies indicate that subjects suffering from depression are far more inclined than non-depressed people to negatively interpret life events and to retrieve and focus on negative memories in order to explain their present trauma or to justify pessimistic predictions.[55] Further, take what Cassam calls one's "intellectual character," that is, peculiarities of one's intellectual constitution—gullibility, carelessness, or inclinations to detect patterns—that can undetectably corrupt one's perspective.[56] Or consider broad psychological pathologies. Take someone who suffers from paranoia—let's call her Susan—who believes she herself can do no wrong, and that anyone who

[55] Sonja Lyubormirsky, Nicole D. Caldwell and Susan Nolen-Hoeksema, "Effects of Ruminative and Distracting Responses to Depressed Mood on Retrieval of Autobiographical Memories," *Journal of Personality and Social Psychology*, (1998) Vol. 75. No. 1, pp. 166-177.

[56] See Quassim Cassam, *Self-Knowledge for Humans*, (Oxford University Press, 2014), pp. 200-201, and "Bad Thinkers," https://aeon.co/essays/the-intellectual-character-of-conspiracy-theorists. Cassam also cites Zagzebski's list of "intellectual vices" such as negligence, idleness, rigidity, obtuseness, prejudice, lack of thoroughness, and insensitivity to detail. See Linda Trinkaus Zegzebski, *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*, (Cambridge University Press, 1996).

criticizes her is "out to get her" and is part of the "conspiracy against her."[57] Clearly, it would be to Susan's advantage to acknowledge and address her paranoia. However, to suggest to Susan that she suffers from paranoia will only trigger the same insecurity that gives rise to her paranoia, and will, as a consequence, cause her to reject the criticism and regard it as just another detail of an orchestrated plot against her. That is to say, for the very reason of her paranoia, Susan will be incapable of recognizing her paranoia.

Finally, some have suggested that the very standpoint of one's cognitive apparatus is conceptually inaccessible. Thus Velleman asserts that "a person can never conceive of his own conceptual capacity from a purely third-personal perspective, because he can conceive of it only with that capacity, and hence from a perspective in which it continues to occupy [a] first-person position."[58] Indeed, Velleman plausibly compares this conceptual impossibility to that of establishing "a visual perspective from which the point between his eyes isn't 'here'".[59] As such, on Velleman's account, facts of one's cognitive or psychological disposition will not only often undermine the accuracy of one's initial self-assessments, but

[57] This example was modeled after Hilary Kornblith's similar case of "Jack," in "What Is It Like To Be Me?" *Australasian Journal of Philosophy*, (March 1998), Vol. 76, No. 1, pp. 48-60, p. 51.

[58] Velleman, "Identity and Identification," from Sarah Buss and Lee Overton, (2002, p. 114). See also Nagel's discussion of one's essential "blind spot": "[H]owever much we expand our objective view of ourselves, something will remain beyond the possibility of explicit acceptance or rejection, because we cannot get entirely outside ourselves, even though we know there is an outside" (1986, p. 128).

[59] "Identity and Identification," from Sarah Buss and Lee Overton, (2002, p. 114).

will undermine one's efforts to recognize the extent to which one's cognitive or psychological dispositions have been compromised.[60]

Now let's consider ways in which our motivations—our *stake* in a matter—commonly prevent us from achieving an unbiased perspective on our existing values, beliefs and desires. Consider Brian, who desperately wants to please his homophobic father, and is therefore incapable of recognizing sexual feelings he has for Todd. Indeed, rather than recognizing such feelings, Brian takes up a militantly anti-gay position and condemns every form of homosexual attraction as a perversion. When a therapist asks Brian to take a step back and more carefully examine his feelings, his fear of his father's judgment automatically kicks in and only reinforces his prejudice. Or consider Jerry and Jack, good friends who have both applied to a prestigious graduate program. Jerry is accepted but Jack is not. Upon learning of his rejection, Jack claims that, in fact, this program is overrated. Jerry suggests that Jack is just upset and jealous at having not been accepted, whereupon Jack "takes a step back" and examines his feelings, but can still detect no evidence of disappointment and jealousy, because, as a matter of fact, he "never really cared so much about that overrated program anyways." As such, Jack has "subjectively optimized" his circumstances.

Not only is our perspective corrupted by such distortions, but note, *crucially*, that it is the very nature of such distorting influences such that they are *designed* to function *beneath the radar* of our conscious awareness and detection. Indeed,

[60] See again Blackburn: "We can compare the situation to looking at our own eyes in a mirror. We might see that our eyes are cloudy; but if they are, it will be with cloudy eyes that we see it." (1998, p. 261) See Blackburn (1998, p. 260) for an excellent treatment of additional non-motivational distorting influences.

as Timothy Wilson has put it, our confidence in our ability to consciously detect and overcome such influences "vastly underestimates the role of non-conscious processing in humans" (2002, p. 21). Wilson asserts that we have developed a "psychological immune system" (2002, p. 38)[61] that protects us from threats to our psychological wellbeing, and this immune system ensures that our biases are working appropriately. The central rule of this system is: "Select, interpret and evaluate information in ways that make me feel good" (2002, p. 39). As such, the effectiveness of non-conscious processing *requires* that it function beneath our conscious radar. For if we suspected that our self-assessments were just the outputs of our "feel-good" subconscious mechanisms, we could no longer take such assessments as seriously, and, in that case, our subconscious mechanisms would have failed to do their job.

And so we can now understand what principally accounts for the inadequacy of the Perceptual Accuracy Hypothesis. Recall again what is involved in overcoming our misperception of the man-standing-at-the-side-of-the-road. When I focus more attention on this object, I more fully engage my perceptual mechanisms, and, as a result, the object in question *passively* submits to such adjustments. That is to say: these objects mount no calculated *resistance* to my corrective efforts. Again, I am not suggesting that, in virtue of my corrective efforts, I will *necessarily* overcome my misperceptions (recall the Muller-Lyer illusion and other such illusions). But the *objects of perception themselves* will not *actively attempt* to sabotage or circumvent my corrective efforts, further

[61] On the psychological immune system, see also Leon Festinger, *A Theory of Cognitive Dissonance*, (Stanford University Press, 1957).

camouflaging themselves or shape-shifting to safeguard their misleading appearance. Yet the same *by no means* can be said when the objects of perception are my *own* features. That is, Korsgaard and Nagel are surely correct in asserting that my *ambition* in taking a reflective backward step is to transform the features of the "me" into objective data on par with mailboxes and traffic cones that will passively submit to the scrutiny of the observing "I". And yet, the "me" that I *take* to be a passive object of perception will mount a calculated and undetectable resistance, ensuring that "I" arrive at a conclusion that is sensitive to *its* needs. Indeed, as Arpaly has put it: "Almost nothing has been written about the agent who steps away from her desires and, as it seems to her, chooses calmly between them, feeling apparent mastery over temptation and emotion, while the very 'I' that steps away from the desires is the unconscious dupe of other desire*s*, emotions, or irrationality."[62] While Arpaly is here speaking specifically of practical reasoning, the same applies to efforts at self-knowledge. That is to say, even when the captainly "I" is convinced it has pulled back and is exercising detached epistemic authority with respect to reading the facts of the "me," the very activity of the "I"'s self-detection is often hijacked by the facts of the "me."

3.3 On Authoring the Page —The Rendering of Facts

I have thus far focused on the question of whether the "I," in virtue of assuming a self-conscious perspective on the "me," can achieve an accurate assessment of the *existing* facts of the "me"—can recognize what is *already there*. And we have seen how features of the "me" can subvert one's investigations. Yet one often

[62] Nomy Arpaly, *Unprincipled Virtue*, (Oxford University Press, 2003), p. 20.

obtains self-knowledge through a different route, that is, through actively "producing" or "rendering" facts about oneself. In these cases the particular fact—usually a belief or desire—in question is not an item one simply recalls or "reads off the page," but something that one actively *authors*.[63] Suppose, for example, I am asked the following question: "Is it right to inform my friend that her husband is cheating on her?" And suppose I haven't considered the matter and therefore hold no belief regarding it. As a result, I must take up a "practical" or "deliberative" stance with respect to the matter: I must "figure out" or "decide" what I believe. And, here too, in the authoring of my belief, self-consciousness is said to play an indispensible role, affording the "I" a set of epistemic privileges and powers.

What specific privileges and powers does the self-conscious "I" wield with respect to the authoring of a belief?[64] One such power is trivially true: that is, *by definition*, a creature lacking self-consciousness could not self-consciously undertake such a process of deliberate belief-formation. That is to say, one often upgrades one's beliefs, or comes to believe new things without a self-conscious

[63] As Moran puts it: "When we speak of 'authority' in connection with first-person statements of belief and other attitudes […] it is not just the report that the person is author of, but also, in a central range of cases, the person can be seen as the author of the state of mind itself, in the sense of being the person who originates it and is responsible for it." (2001, p. 113)

[64] Here I will bracket a broader question, that is, to what extent is self-consciousness necessary for belief-formation in general? That is to say, in order for an organism to hold beliefs in general, must the organism be capable of *knowing* that it holds at least some beliefs? For an interesting discussion of this question, see Hilary Kornblith, *On Reflection*, (Oxford University Press, 20012) pp. 55-61.

awareness that one is doing so.[65] But one cannot voluntarily and self-consciously author a belief without voluntarily and self-consciously doing so. Indeed, this process of self-consciously authoring beliefs, if anything, argues in support of the view that one can "take up of the reins" of one's reflective activity, that one can actively "conduct" it. Thus, just as in the previous section we allowed for the fact that self-consciousness renders us *capable* of knowing that we possess certain facts (even if our self-conscious analysis of such facts does not ensure an accurate reading of them), in this case we must at least grant that self-consciousness renders us *capable* of voluntarily engaging in the process of producing certain attitudes. What is left to be determined is the question of what, if any, special powers the "I" can bring to this process.[66]

The Rationalist suggests the following: The unique role that the "I" plays in one's deliberative process is that of a "rational agent." That is, for example, when I am asked, "What do you believe about p?" and I hold no belief about p, I take the question to be asking, "What I *should* believe about p?" and I take *this* question to be asking, "What is *most rational* to believe about p?" It is in my efforts to arrive at the most rational belief about p, both with respect to theoretical and practical questions, that the "I" is uniquely positioned to appeal to the "authority of reason". What renders the "I" uniquely positioned to engage in such

[65] That this is the case can be seen for the following reasons. First, if every change or upgrade in belief availed itself to consciousness, our consciousness would be overrun with a cascade of beliefs. Second, such a requirement on belief would generate a regress, e.g., I can only believe it is raining if I am conscious of the fact that I believe it; I can only believe that I am conscious of my consciousness of my belief if I am conscious of that belief, and so on.

[66] The separate question whether, or to what extent, the "I" can truly exercise such *agency* will be taken up in the following chapter.

activity? To arrive at an adequate response to this question we will need to more deeply explore nature of rational agency.[67]

3.3.1 On the Conditions of Rational Agency

The question is: What renders the self-reflecting "I" uniquely suited to engage in rational agency? For a preliminary understanding, let us first turn to Kant. Recall Kant's assertion that: "The fact that the human being can have the "I" in his representations [i.e., is self-conscious] raises him infinitely above all other living beings on earth." Now, Kant also tells us, "a human being really finds in himself a capacity by which he distinguishes himself from all other things, even from himself insofar as he is affected by objects, and that is reason" (*Groundwork*: 4:452)[68]. So both the capacity for self-representation—for representing himself as an "I"—and the capacity for reason, render a human being superior to all other living things. Yet our capacity for reason possesses a further distinguishing feature. For, Kant tells us, "Only a rational being has the capacity to act in accordance with the representation of laws, that is, in accordance with principles, or has a will" (*Groundwork*: 4:413). That is, whereas unknown "laws" immediately regulate the behavior of irrational creatures, a human being can both *represent* such laws, and can *formulate* laws or principles according to which she can regulate her behavior. So, putting these pieces together: We human beings are distinguished from all other living things in virtue of our capacities to represent

[67] Note again that the following analysis will be further developed in the following chapter.

[68] The question of why a person's capacity for reason distinguishes him "even from himself" will be explored in Chapter 6.

ourselves, to represent laws or principles, and to govern our behavior in light of these laws or principles, as dictated by reason. That is, in light of such laws or principles, we can infer various consequences of possible choices, run cost-benefit analyses, or subject various proposals to the Kantian "universalizability" test[69].

And yet, according to Kant, at least when it comes to practical questions, a good deal more is required of us if we wish to successfully engage our rational faculties. For Kant provides that when we deliberate we "cannot act otherwise than under the idea of freedom".[70] This "idea of freedom" should be understood both in descriptive and normative terms. In the first case, as a matter of descriptive fact, when we deliberate, we cannot help but believe that the very act of deliberation is up to us. In the second, normative, sense, when we deliberate we feel obligated to make certain that the course and outcome of our deliberations is not determined by, as Kant puts it, *heteronymous* forces—forces arising from an *external*, or *alien*, source—but that it they up to our *rational will*. As Kant puts it: "*Will* is a kind of causality of living beings insofar as they are rational, and *freedom* would be that property of such a causality that it can be efficient independent of alien causes *determining* it" (1997a: 4:446, emphasis his). That is, insofar as we desire that we, by means of our rational will, determine the course and outcome of our deliberations, we must make certain that our reflections are

[69] That is, we can undergo that decision-making procedure whereby we determine whether or not we can will our guiding maxim as universal law. See Kant, *Groundwork of the Metaphysics of Morals*, (ed.) Mary Gregor), (intro.) Christine Korsgaard, (Cambridge, 1997a), p. 31, 4:421.

[70] See 1997a, 4:448: "[E]very being that cannot act otherwise than **under the idea of freedom** is just because of that really free in a practical respect, that is, all laws that are inseparably bound up with freedom hold for him just as if his will had been validly pronounced free." (emphasis his)

governed by our reason, for "unless reason holds the reins of government in his own hands, a human being's feelings and inclinations play the master over him."[71] As such, our self-regulating, self-legislating rational will permits us to overcome the influence of heteronymous impulses that might otherwise dictate the outcome of our deliberations. In this way, we can make certain that *w*e—our rational wills—are making up our own minds.

We have already encountered this general idea in Korsgaard, who assures us that our deliberations should not, and, indeed, *cannot*[72], be determined by our "impulses." Rather, our impulses just provide data we *feed* to our rational faculties, data we may or may not regard as germane to our deliberative conclusions. And yet some Rationalists propose that we must go *even further* if we wish to successfully engage our superior rational faculty. Indeed, as Moran puts it: "In deliberating about some matter I do not even take as fixed whatever stock of beliefs and desires I may bring to the problem, for entering into the spirit of rational deliberation means that I acknowledge that reflection on the problem may lead me to abandon or revise any one of them" (2001, p. 133). That is,

[71] Kant, *The Metaphysics of Morals*, (trans. & ed.) Mary Gregor, (intro.) Roger J. Sullivan (Cambridge, 1996) 6:408.

[72] Korsgaard expresses this "cannot" as a "must." She writes: "Once the space of awareness—as reflective distance, as I like to call it—opens up between the potential ground of a belief and the belief itself, or between the potential ground of an action and the action itself, we *must* step across that distance with some awareness that we are doing so, and so *must* be able to endorse the operation of that ground as a basis for what we believe or do. This means that the space of reflective distance presents us with both the possibility and the necessity of exerting a kind of control over our beliefs […] And it is the same fact that we now both can have, and absolutely require, reason to believe and act as we do" (2008, p. 4-5, emphasis mine). See also Nagel: "Having stopped the direct operation of impulse by interposing the possibility of decision, one can get one's beliefs and actions into motion again *only* by thinking about what, in light of the circumstances, one should do" (1997, p. 109, emphasis mine).

according to Moran, when one deliberates on some matter, one must not regard *any* of one's standing beliefs as immune to revision. For, if one were to do so, then one would "[take] it to be an open question whether this activity will determine what one actually does or believes" (2001, p. 133). As such, in order to exercise full-fledged rational autonomy, an agent must believe that (1) He is the conductor of his deliberations; (2) He can determine the course and outcome of his deliberations exclusively by means of his rational will; and (3) He can bring a sufficiently open and unbiased mind to his deliberations, such that his deliberations will not be undermined by opaque, or unquestionable, beliefs.

One can appreciate the appeal of such a portrait of an ideal rational agent—the *Homo Philosophicus*, as Cassam calls her[73]. Indeed, there are few insults more stinging than that of being labeled "irrational," or "unreasonable," or, even worse, unknowingly "biased" or "prejudiced." Further, our capacity for competent, efficient deliberation seems uniquely to certify our elevated human status, our superiority to the rest of the "irrational" living kingdom.[74] Yet one may wonder how realistic such a portrait really is. For recall that our review of the literature on self-perception showed that even when subjects of studies were informed of their biases and were asked to review their beliefs in light of this fact, they routinely insisted upon their previously-stated opinions, denying that they were thus biased. And recall the vast array of non-motivational and motivational influences that undetectably undermine our critical faculties—influences, indeed, that are *designed precisely to undermine us in ways that elude our detection*. As such, if

[73] Cassam (2014).
[74] Recall Kant's "infinite superiority" claims and see Aristotle's *Function* argument. (2002, 1097b-1098a20).

our examinations of our *existing* belief states are susceptible of such corrupting influences, it stands to reason that these distorting influences will just as agressively undermine our efforts to rationally produce or author our conclusions, attitudes and actions. Indeed, given the complexity of the deliberative process, the demands involved in the examination of evidence, the variety and sophistication of inferences that one must successfully perform, it would seem *much more likely* that such influences would exert themselves.[75]

Now, I am not suggesting that a human being *cannot possibly* achieve the status of Homo Philosophicus. And I am not contending that the taking up of the deliberative stance *does not* or *cannot* render us capable of appealing to the authority of reason. It certainly can and does, however problematically. A further question is whether or not the success of my rational engagement should be understood as a kind of *achievement*—as some feat or accomplishment for which "I" am genuinely *responsible*. That question will be a central concern in the following chapter. Before moving on, however, let us explore one further way we can establish facts about ourselves—that is, through acts of "self-identification" or "self-constitution".

3.4 On Identification

Thus far we have considered two means whereby the assumption of a self-conscious perspective—i.e., the taking of the "backward step"—ostensibly

[75] Indeed, even Kant admits that "it is absolutely impossible by means of experience to make out with complete certainty a single case in which the aim of an action otherwise in conformity with duty rested simply in moral grounds" (1997a: 4:407) See also Kornblith, "Distrusting Reason" Midwest Studies in Philosophy, XXIII (1999).

permits the "I" to establish the facts of the "me". In the first case, the backward step was supposed to *objectify* my facts and thereby render them more susceptible of accurate detection. In the second, it was proposed that the "I," in virtue of its ability to detach from personal impulses, could appeal to and engage a *superior rational force* in the service of authoring beliefs and desires. I will now explore a third means whereby, on certain Rationalist accounts, the "I" can gain epistemic purchase on the facts of the "me," that is, through acts of *identification*. That is, according to such thinkers, the very act of "endorsing," "committing to" or "deciding in favor of" certain desires or beliefs *renders it the case* that such facts are more self-representative, or more "mine," than they otherwise would be. I will begin with Frankfurt's discussion of identification and then move on to Korsgaard's account. I will then discuss Schechtman's related views on "narrativity" and Moran's notion of "avowals." In each case, I will first briefly present the view and then I will offer commentary.

3.4.1 Frankfurt

Among contemporary philosophers perhaps Harry Frankfurt has been most instrumental in introducing and developing the notion of "identification".[76] His view is rather complex, so we will need to lay some groundwork. First, Frankfurt tells us that "[t]o be a person, as distinct from simply a human organism, requires

[76] Frankfurt's views on these matters have changed over time. I will address his earlier views in this chapter. I will more fully explore his latter views in later chapters.

a complex volitional structure involving reflective self-evaluation."[77] This volitional structure consists in first-order desires (I desire a vodka-tonic); second-order desires, that is, desires about first-order desires (I desire or don't desire to have the desire for the vodka-tonic), and second-order volitions, that is, a second-order desire that a first-order desire carries one to action. Now, when one evaluates one's first-order desires (to drink the vodka tonic), and develops a second-order desire (not to have this desire because one is an alcoholic and knows one must abstain from alcohol) and then advocates for this second-order desire, one is "identifying" with this second-order desire (and, in so doing, further identifying with or else detaching form the first-order desire). Such identifications are expressive of one taking an "active" role with respect to a particular desire or action, rather than occupying the position of a mere "passive bystander" with respect it.[78] As such, in endorsing or identifying with a particular desire, one assumes *responsibility* for it and thereby "makes one of [his desires] more truly his own".[79] Likewise, as mentioned, one can "withdraw from,"

---

[77] "The Faintest Passion," from *Necessity, Volition, and Love*, (Cambridge University Press, 1999) p. 103 FN.

[78] Indeed, Frankfurt claims that when it comes to second-order volitions, "it is impossible for [a person] to be a passive bystander to them. They *constitute* his activity—i.e., his being active rather than passive—and the question of whether or not he identifies himself with them cannot arise." See "Three Concepts of Free Action," from *The Importance of What We Care About*, (Cambridge University Press, 1988) p. 59-60 (emphasis his).

[79] Frankfurt, "Freedom of the Will" from 1988, p. 18. Frankfurt elaborates: "The pertinent desire is no longer in any way external to him. It is not a desire that he 'has' merely as a subject in whose history it happens to occur, as a person may 'have' an involuntary spasm that happens to occur in the history of his body. It comes to be a desire that is incorporated into him by virtue of the fact that he has it *by his own will*. […] Through his action in deciding, he is responsible for the fact that the desire has become his own in a way in which it was not unequivocally his own before." "Identity and Wholeheartedness," from 1988, p.

"externalize," or "alienate," a desire, in such a way as to render this desire *no longer* his own (or, at least, *less truly* his own). Thus, Frankfurt tells us: "It is these acts of ordering and rejection—integration and separation—that create a self out of the raw materials of inner life. They define the intrapsychic constraints and boundaries with respect to which a person's autonomy may be threatened even by his own desire" (1988, p. 170). As such, according to Frankfurt, when we undertake this process of "ordering and rejection" we are literally *creating* our identities, determining which desires or beliefs represent us. And this process is of paramount importance, for, as he says, "it is a salient characteristic of human beings, one which affects our lives in deep and innumerable ways, that we care about what we are" (1988, p. 163).

Such is Frankfurt's view. And I will allow that it *does* track our phenomenology. After all, we sometimes *do* review our spectrum of attributes and wish to embrace, be moved and defined by, some attributes, while wish to reject or *not* be moved or defined by, others. I may wish, for example, that my desire to be helpful or generous were more self-representative than my selfish desire to serve myself, or that my desire to be accepting and tolerant was more self-representative than my tendency to be judgmental or confrontational. And we believe that such acts of endorsement (identification) or rejection (externalization) should make a difference with respect to the kind of people we really are.

---

170. See also: "A person who cares about something is, as it were, invested in it. By caring about it, he makes himself susceptible to benefits and vulnerable to losses depending upon whether what he cares about flourishes or is diminished. We may say that in this sense he identifies himself with what he cares about" (From, "On the Necessity of Ideals," 1999, p. 111).

While Frankfurt's analysis tracks our hopeful intuitions, however, it is doubtful whether our optimism is justified. Notice, first, that the identification with a desire, however active, by no means guarantees the self-representational authority of that desire; nor should the externalization of a first-order desire necessarily certify its non-representational status.[80] Let us call this the *Authentication Problem*. So, consider Brian who discovers in his adolescence that he is attracted to Todd. Due to his horror of such desires, or his horror of his father's disapproval, Brian develops a second-order desire *not* to possess his first-order desire. In this case, Todd will identify with his second-order desire and dis-identify with or alienate his first-order desire. But clearly Brian's identification with the second-order desire does not render it *more* self-representing; nor does it render his first-order desire *less* self-representing.

Yet a deeper problem lurks. For if a belief or desire must depend on a higher-order belief or desire to vouch for its self-representational authority, a regress will ensue—each desire requiring the certifying authority of a higher-order desire, ad infinitum. To halt such a regress, one will need to invoke a desire one has not endorsed, and yet, in virtue of its lack of endorsement—given Frankfurt's conditions on self-representational authority—nothing will vouch for its authority, and this, in turn, will threaten to delegitimize the self-representational status of the entire chain of desires.

Responding to these concerns, Frankfurt suggests that one can both terminate a regress and certify the representational authority of a desire if one can make a

[80] Frankfurt concedes this point. As he puts it: "I have maintained that the question of whether a passion is internal or external to a person is not just a matter of the person's attitude toward the passion." (1998, p. 64).

"decisive commitment" to it. Under such conditions, as he puts it, the commitment will "[resound] throughout an endless array of higher orders" (1998, p. 21). But what, precisely, is so special about the decisive nature of such a commitment? That is, as Watson has put the problem: "What gives [a decisive commitment] any special relation to 'oneself'? It is unhelpful to answer that one makes a 'decisive commitment,' where this just means that an interminable ascent to higher order is not going to be permitted."[81] Frankfurt suggests that what gives such a decisive commitment a special relation to the self is that one can make a *wholehearted* commitment to it; that is to say, one can make such a commitment *without reservation*. Indeed, Frankfurt compares the process of wholeheartedly identifying with a particular desire with the process of performing arithmetic calculations. He says:

> [B]oth in the case of desires and in the case of arithmetic a person can without arbitrariness terminate a potentially endless sequence of evaluations when he finds that there is no disturbing conflict, either between results already obtained or between a result already obtained and one he might reasonably expect to obtain if the sequence were to continue. Terminating the sequence at that point—the point at which there is no conflict or doubt—is not arbitrary. For the only reason to continue the sequence would be to cope with the actual conflict or with the possibility that a conflict might occur. Given that the person does not have this reason to continue it is hardly arbitrary for him to stop. (1998, p. 169)

Frankfurt's suggestion, thus, is that an agent's decision carries special or non-arbitrary authority just when that agent, in making his decision, is *met with no resistance*. Yet Frankfurt's further refinement of the conditions for authentic identification still fails to adequately address Watkins' concern. For, indeed, what

[81] Gary Watson, "Free Agency" *The Journal of Philosophy*, Vol. LXXII, No. 8, April 24, 1975, p. 218.

prevents a person from being misguided even here, that is, even with respect to the "wholeheartedness" of his decision? Recall the aforementioned bevy of empirical evidence testifying to our tendencies toward self-misrepresentation. As such, there seems nothing in principle to prevent one from misjudging oneself even with respect to the question of one's wholehearted commitment.[82]

3.4.2 Korsgaard

Let us now examine Korsgaard's Frankfurt-friendly take on the process of identification and see whether she can overcome the Authentication Problem. Like Frankfurt, Korsgaard believes that all persons are continuously engaged in a project of constructing and refining their "practical identity," that is, "a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking" (1996b, 101). Moreover, Korsgaard asserts, with Frankfurt, that what renders a particular feature "mine" is my *active* endorsement of, or identification with, a belief, desire, or principle. By what means does one actively endorse or identify with a particular belief or desire? I quote the following lengthy passage in full:

> A rational will is a self-conscious causality, and a self-conscious causality is aware of itself as a cause. To be aware of yourself as a cause is to identify yourself with something in the scenario that gives rise to the action, and this must be the principle of choice.

[82] Indeed, Frankfurt concedes this as well. He tells us: "Indeterminacy in the life of a real person cannot be overcome by preemptive decree. To be sure, a person may attempt to resolve his ambivalence by deciding to adhere unequivocally to one of his alternatives rather than to the other; and he may believe that in thus making up his mind he has eliminated the division in his will and become wholehearted. Whether such changes have actually occurred, however, is another matter." "The Faintest Passion," from (1998, p. 101). Again, Frankfurt will refine these views in his later work that I will address in Chapter 5.

> […] You have some principle which favors *A* over *B*, so you exercise this principle a, and you choose to do *A*. In this kind of case, you do not regard yourself as a mere passive spectator to the battle between A and B. You regard the choice as yours, as the product of your own activity, because you regard the principle of choice as expressive, or representative, of yourself. You must do so, for the only alternative to identifying with the principle of choice is regarding the principle of choice as some third thing in you, another force on par with the incentives to do A and to B, which happened to throw its weight in favor of A, in a battle at which you were, after all, a mere passive spectator. But then you are not the cause of the action. Self-conscious rational agency, then, requires identification with the principle of choice on which you act. (1999, p. 26, bold mine)

To grasp what Korsgaard is getting at here, it will be necessary to quickly return to Kant. Recall that, on Kant's account, one must deliberate under the "idea of freedom." That is, from a practical standpoint, the self-conscious rational agent cannot help but regard herself as the active *cause* or *conductor* of her own reflective process, and, as such, as the *producer* or *author* of her practical conclusions. These are not activities toward which a person can stand as a Frankfurtian "passive bystander"; rather, they are actions that one must self-consciously *conduct*. Now, to the extent that an agent cannot help but view herself as such, she must likewise regard her reflections, and her ultimate endorsement of a particular principle for action, as something that has been *up to* her, as something for which she is *responsible*.[83] As such, Korsgaard proclaims, "Self-conscious rational agency, then, requires identification with the principle of choice on which you act." And yet it is unclear how Korsgaard's analysis can address the Authentication Problem, that is, the question: *By what means* does the active identification with a particular desire or principle *render it the case* that this

[83] Again, I will be developing these ideas at greater length in the next chapter.

desire or principle is authoritatively self-representating? Or, to put it a little differently: In virtue of *what* does one's active identification with a particular principle or attitude *render* this principle or attitude "more one's own" than a principle or attitude that she has not endorsed?

To understand Korskaard's curious response to this question, we will need to step further back and acquaint ourselves with some more of her conceptual machinery. Note, first, that, on Korsgaard's view, while irrational animals are capable of mere "behavior" or "movements," a *person* is capable of "action," and, "[w]hat distinguishes action from mere behavior and other physical movements is that it is *authored* – it is in a quite special way attributable to the *person* who does it, by which I mean, the *whole* person" (Korsgaard, 1999, p. 3). What renders a person "whole"? Korsgaard explains that what renders a person, or any organized entity, whole, is the property of being *united*. As she puts it: "The actions which are most truly a person's own are precisely those actions which most fully unify her."[84] Thus a person is whole insofar as her parts—her impulses, desires, and beliefs—are unified.[85] Moreoveer, following Kant, Korsgaard asserts that one's parts can *only* be unified by means of one's autonomous, authoritative, rational, will. Only in this way can the *person*, and not heteronymous elements working *in* her or *on* her, draw her together and determine her behavior. As such, Korsgaard tells us:

[84] Korsgaard's derives her conception of "unity" partly by appeal to Plato's discussion of the ideal city in *The Republic* (see especially 443d-e). Here, Plato argues that a city will not be able to properly function if its parts are not united under the authority of its rulers.

[85] See Kant's similar view of "self-mastery" which involves "bring[ing] all his capacities and inclinations under his (reason's) control and so to rule over himself" (*MM* 6:408).

> The actions which are most truly a person's own are precisely those actions which most fully unify her and therefore most fully constitute her as their author. They are those actions which both issue from, and give her, the kind of volitional unity which she must have if we are to attribute the action to her as a whole person. (1999, p. 3)[86]

Thus, on Korsgaard account, in order for a human being to be capable of action, her parts must be united, and the very act of deliberation, of organizing oneself according to a principle, *just consists* in the gathering together and the uniting of one's parts. The very act of deliberation thus *constitutes* one's personhood, and, insofar as this is the case, the "[b]eliefs and desires you have actively arrived at are more truly your own than those which have simple arisen in you (or happen to inhere in a metaphysical entity that is you)."[87]

3.4.2.1 Critique of Korsgaard

So, let us examine what I take to be two of Korsgaard's most critical and related claims. One is that the agent must identify with her principle of choice because she must view her principle of choice, and indeed every element of her deliberative process, as being active, or up to her. The second claim is that, in virtue of the fact that the act of deliberation constitutes the self by means of uniting its "parts,"—in virtue of this fact, the very *determination* of a particular principle as self-representing, will somehow *render* that very fact more genuinely

[86] See also: "When you deliberate about what to do and then do it, what you are doing is organizing your appetite, reason and spirit, into the unified system that yields an action that can be attributed to you as a person. Deliberative action pulls the parts of the soul together into a unified system. (1999, p. 22)

[87] Korsgaard, *Creating the Kingdom of Ends*, (Cambridge: Cambridge University Press, 1996), p. 379. She says elsewhere: "[A] desire or belief that has simply arisen in you may be reflectively endorsed, and this makes it, in the present sense, more authentically your own." (1996) p. 394, note 34.

self-representating, i.e., "more truly one's own" than it otherwise would have been. Let us consider some objections to these two claims.

In the first place, it should be noted that the self-conscious deliberative procedure Korsgaard recommends is certainly not a necessary condition for arriving at a rational, self-representing principle. Consider, as Arpaly has, instances of "dawning"—sudden epiphanies or realizations that come as though involuntarily, but strike us as perfectly rational and representative of our true thoughts or feelings. One, for example, may "suddenly realize" that one can no longer tolerate one's job; or it may suddenly dawn on one that one loves one's acquaintance. Indeed, as Arpaly puts it, "Dawning processes are perhaps the main way in which people change their minds, especially concerning subjects they regard as important—the very subjects regarding which an attempt to argue with them and talk them out of the error of their ways is likely to encounter the sternest irrational resistance."[88] And yet we do not regard the spontaneous, seemingly involuntary nature of such a dawning phenomena as undermining its self-representational authority.

Second, there seems nothing in Korsgaard's account that renders it superior to Frankfurt's with regard to guaranteeing the self-representational authority of one's conclusions. For, indeed, we may grant much of what Korsgaard says—that is, we may grant that it is by means of the deliberative procedure that we consciously attempt to "pull ourselves together," or "unify our parts." But again, there seems to be nothing that safeguards our deliberative *activity* from being an

[88] Arpaly (2003) p. 55. See also Nisbett and Wilson (1977), p. 240-241 who offer a nice review of the mysterious process whereby artists and scientists involuntarily arrive at solutions to problems.

exercise in fantasy or wish fulfillment. Indeed, consider, as Arpaly does, Mark Twain's Huckleberry Finn deciding not to turn Jim—a black slave—in to Miss Watson, Jim's owner. At one point Huck decides that it would be best to turn Jim in, that such a decision would be most consistent with what his "conscience" tells him, and thus would best express who he *truly* is. Yet when Huck has the opportunity to turn Jim in, he fails, attributing his failure to his weak will, claiming he doesn't have "the spunk of a rabbit." Again, Huck has arrived at his conclusion, at his chosen behavior, through a painful, self-conscious, deliberative procedure. But Huck's failure stems at least in part from the fact that his *considered opinion*—the opinion that seemed most rational and most self-representative—in fact *does not* reflect his true desires, and hence is not authoritatively self-representing. For, unbeknownst to himself, over the course of Huck's travels with Jim, Huck has developed feelings—among them, respect—for Jim, and *such* feelings actually express Huck's true nature. Thus Huck may have endeavored to master certain impulses in order to "unify" himself and to seize upon some principle that is genuinely self-identifying. But his efforts neither appeared to conduce to self-unity (after all, he ultimately "broke down"), nor to identify a correctly self-identifying principle.

3.4.3 Schechtman and Moran

Frankfurt's and Korsgaard's accounts of identification place special emphasis on the role of *endorsement*. We seize upon a particular attitude; we desire that such an attitude be self-representative, and we *render* such an attitude "mine" or

"more mine" by means of our active endorsement and identification.[89] But we have seen that, in spite of such acts of decisive endorsement, we may be mistaken as to whether a belief or desire is genuinely self-representing. Yet there are other routes one can take to establish one's ownership of a particular desire or belief—routes which may be more assured. On these accounts, identification will lay *not* in the intellectual endorsement of an attitude, but rather in the deeper *recognition* or *assimilation* of an attitude. Let us therefore briefly consider such views as developed by Schechtman and Moran.

On Marya Schechtman's "narrative" account, "a person creates his identity by forming an autobiographical narrative—a story of his life."[90] Here again, a person actively contributes to or shapes his or her self-conception, and this shaping is narrative in form "insofar as the incidents and experiences that make up his life are not viewed in isolation, but interpreted as part of the ongoing story that gives them significance" (1996, p. 97). Now, Schechtman does not suggest that a person can make up one's story arbitrarily or whole cloth; rather, one's story must respect a set of "reality constraints"—both observational and interpretive, such

[89] Note that both Frankfurt and Korsgaard's accounts allow for *degrees* of endorsement. As I pointed out above (note 78), Frankfurt says as much, conceding that there can always be an element of indecisiveness or ambivalence in one's commitment. Korsgaard suggests the same: "Since some ways of acting unify their agents better than others, the extent to which a movement is an action is a matter of degree: some actions are more genuinely actions than others (2008), p. 45. See also her discussion of "Standards for Action" (1999), p. 12-15.

[90] See Marya Schechtman, *The Constitution of Selve*s (Cornell University Press, 1996), p. 93. For another excellent treatment of the necessity of both implicit and explicit storytelling, see Damasio (2010), esp. 311: "Implicit storytelling has created our selves, and it should be no surprise that it pervades the entire fabric of human societies and cultures." For a strong objection to the necessity of narrativity for the constitution of personhood, see Galen Strawson, "Against Narrativity," *Ratio (new series)* XVII 4 December 2004, pp. 428-452.

that, as she puts it: one's story must "fundamentally cohere with reality" (1996, p. 119).[91] Yet, in the authoring of one's story, Schechtman claims that the *very recognition and articulation* of one's particular features renders these components more "hers" than those elements she has not yet recognized. As she puts it: "The elements of a person's narrative he cannot articulate are his, but […] they are only partially his—attributable to him to a lesser degree than those aspects of the narrative he can articulate" (1996, p. 117). Thus, consider again Brian who may have very little knowledge of the fact that he is romantically desirous of Todd. According to Schechtman, the more Brian consciously grasps the fact of his desire and the more clearly he can discern its features and work them into his consciously constructed narrative, the more he will *own* this fact, and, thus, the more self-representational this fact, and its attendant desires, will be. In virtue of what does Brian's recognition and articulation render his desire more self representative? Schechtman plausibly argues that a person's unrecognized and unarticulated features will influence him in an impulsive, mysterious, rigid and automatic fashion, whereas once one better recognizes these attitudes one can better tame or control them.[92]

Finally, let us consider Richard Moran's concept of "avowals". Moran contrasts a "theoretical"/"descriptive" stance one can take toward oneself with a "first-personal"/ "practical" stance one can assume. To better highlight the

[91] Indeed, on Schechtman's account, such constraints even include the *style* of one's self-narrative. A self-narrative written in a style "wildly different from those standard in our culture—for example, a self-conception that is not even narrative in form [would not be considered] identity-constituting at all" (1996, p. 105). I will more fully explore the nature of these constraints in Chapter 5.

[92] Schechtman (1996, p. 118).

differences between these stances, Moran casts them in a therapeutic context. A person, for example, may discuss her case with her analyst, and achieve a spectator's theoretical grasp of a number of facts about herself. But, Moran asks: "[w]hat would be missing from a restoration of self-knowledge that remained theoretical and descriptive in this sense?" (2001, p. 89) Moran suggests that any such approach would

> neglect, at the very least, the crucial therapeutic difference between the merely 'intellectual' acceptance of an interpretation, which will itself normally be seen as a form of resistance, and the process of working through that leads to a fully internalized acknowledgement of some attitude which makes a felt difference to the rest of the analysand's mental life. The goal of treatment […] requires that the attitude in question be knowable by the person, not through a process of theoretical self-interpretation but by avowal of how one thinks and feels. (2001, p. 90)

The suggestion here is that a patient may intellectually recognize a particular belief or desire as one she has, but insofar as her grasp of this belief or desire is "only theoretical," it will remain impersonal or detached. Not until the patient can emotionally "own up to" the attitude in question can she fully assimilate or integrate it. To better see this, consider a patient—we will call her Martha—who has been informed of her childhood abuse at the hands of her father. Moran suggests that Martha "may not doubt this. But without her capacity to endorse or withhold endorsement from that attitude, and without the exercise of that capacity making a difference to what she feels, this information may as well be about some other person, or about voices in her head" (2001, p.93). Here again, note Martha *already* possesses an intellectual grasp of the fact of her abuse; she does not doubt this fact; but she has not *emotionally* accepted it such that her understanding will

make a "felt difference". Only once she has done so will she achieve a "fully internalized acknowledgement." Here again, note that what distinguishes Moran's notion of endorsement from Frankfurt's and Korgaard's is that, in Martha's case, she *already* intellectually recognizes and affirms a fact about herself, but it is the *emotional* contribution that brings about her fully internalized acknowledgement. Only then can she *fully* own it and thereby fully psychologically integrate this fact.

3.4.3.1 Critique of Schechtman and Moran

In discussing the views of Frankfurt and Korsgaard, one concern was that one might misidentify one's authentic first-order or higher-order desire. After all, on their accounts how one wishes to identify may well express a form or fantasy or wish-fulfillment. What recommends the accounts of Schechtman and Moran is that they appear to dramatically reduce this possibility. That is, on Schechtman's account, one has already identified a personal feature or attitude that truly is self-representing, and one is just *coming to more fully recognize* or understand this feature or attitude, such that it can be more fully self-integrated. Similarly, on Moran's account, one already intellectually recognizes and identifies with a particular fact but has yet to *emotionally accept* or *own up* to it, and here, again, the act of emotional assimilation facilitates a more thorough integration of the attitude in question. Now, one may contend there is room for error even here. After all, Brian, for example, may more fully understand and emotionally inhabit his self-imposed prejudice against homosexuality, all in an effort to suppress or

hide from his true desires. As such, to more fully recognize and inhabit his prejudice would seem to further alienate him from his true and hence more self-representative desires.[93] Yet it is certainly possible that one *can* at times grasp one's genuine, self-representative facts, and that such deeper acknowledgment *can indeed* make the kind of difference these authors suggest. After all, as discussed above, we are not completely blind to ourselves and we can sometimes achieve a better grasp of our attitudes through a deeper understanding or through an emotional acknowledgment of them.

3.5 Conclusion

While I have argued in this chapter for a set of negative epistemic claims, it should be clear that I am not endorsing any sort of radical skepticism. As opposed to Nietzsche and Camus, I believe that the "I" can attain a more accurate assessment of the facts of the "me." While this may not take place as regularly as we might like, it certainly *can* take place. Likewise, I have not ruled out the possibility that the "I" can harness the "authority of reason" in its pursuit of unprejudiced beliefs and desires—even though, again, the rational fidelity of such pursuits is commonly corrupted by unrecognized influences. I have also allowed that the "I" can identify with a belief or desire, as Korsgaard and Frankfurt have suggested, even if the "I" is often mistaken about the self-representational status of the belief or desire in question. Finally, I have accepted that the "I" can

[93] Schechtman and Moran may contend here that Brian could not *authentically* recognize such a prejudice. I will explore this possibility in Chapter 5.

recognize or avow the facts of the "me" in such a way as to facilitate a more complete integration of such facts.

In all of these respects, the "I" can indeed make at least some headway, however limited and however hobbled, by the epistemic resources at its disposal. A further question is whether any of these operations, their success or failure, is genuinely attributable to, or "up to," the agential "I," or whether these phenomena should rather be understood as further expressions of the facts of the "me." We will explore this question in the following chapter.

CHAPTER 4

AGENCY

4.1 Introduction

In the previous chapter I examined a number of proposals suggesting that the "I" can exercise a set of *epistemic* powers with respect to the "me". The general idea was that the "I," in virtue of its distance from the facts of the "me," could detect, determine, or enhance one's understanding or assimilation of one's facts. I argued that inasmuch as such powers are available to us, they are susceptible to the distorting influence of the facts of the "me", and that, as Arpaly puts it, the "I" often acts as an "unconscious dupe" of such intrusive influences. Now, I believe it is safe to assume that none of us would ever wish to behave as an "unconscious dupe"—indeed, that the very prospect of behaving as such should induce a wave of panic. After all, as Guignon has asserted, one condition on being a "bounded, masterful self," is the capacity to captain one's own crew, to assume ownership of and bear responsibility for one's behaviors and one's attitudes, or, collectively, one's "facts". But what exactly might it mean to do this? That is to say, what might it mean for a person to author, to own, or otherwise command one's facts, as opposed to being a mere *expression* or *composite* of one's facts? This question will be the subject of this chapter. That is, the subject of this chapter will center on the question of what it might mean to function as a responsible agent.

4.2 Minimal or Shallow Vs. Deep Agency

To get started, let us distinguish between two sorts of agency—"minimal" or "shallow" agency and "deep" agency. With respect to the former, I will join the philosophical consensus in understanding shallow agency as a property that distinguishes "actions" (behaviors that are "up to" an entity—e.g., a man *jumps* from here to there), from mere passive "movements" that "happen to" an entity[94] (e.g., a gust of wind *blows* a man from here to there).[95] In the former case, as Aristotle tells us, the "origin of [one's] moving"—or, as I shall call it, the "moving principle"—is *internal* to, as opposed to *external* to, the organism[96]. Take, for example, a sunflower. I may grab hold of the sunflower and turn it in the direction of the sun; or else the sunflower *itself* will turn in the direction of the

[94] I am using the term "entity" here because, per my characterization of minimal agency, I see no impediment to inanimate objects exhibiting minimal or shallow agency. An alarm clock, for example, will ring at a certain hour, and the principle of movement involved in the ringing will be endemic to the alarm mechanism. As such, it appears quite natural to me to attribute minimal agency to mechanisms like alarm clocks.

[95] As Frankfurt puts it, agency comes down to the question of "the difference between passivity and activity." He elaborates: The "difference between passivity and activity is at the heart of the fact that we exist as selves and agents and not merely as locales in which certain events happen to occur" (1998, ix). See Wittgenstein's famous characterization of this problem as one that captures the distinction between my hand "going up" and my "lifting" my hand. See Wittgenstein (2001) §611-630. See also, Korsgaard: "[A]n action requires an agent, someone to whom we attribute the movement in question to its author" (2009, p. 18). And "The authoredness of it—is the essence of action." (2009, p. 83). For a nice survey of these accounts, see, for example, John Hyman and Helen Steward (eds.), *Agency and Action,* (Cambridge University Press, 2004); James Stacey Taylor (ed.) *Personal Autonomy*, *New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy,* (Cambridge University Press, 2005); and Buss and Overton (2002).

[96] See Aristotle, *Nicomachean Ethics* (trans.) Christopher Row and Sarah Broadie, (Oxford: Oxford University Press, 2002): "And a person acts voluntarily in the cases in question; for in fact in actions of this sort the origin of his moving the instrumental parts is in himself, and if the origin of something is in himself, it depends on himself whether he does that thing or not." (1110a15-20).

sun. In the first case, it is clear that I—something external to the sunflower—am the origin of the sunflower's movement, and, as such, that the principle of motion is external to the sunflower. In the latter case, the origin of movement—the network of dynamic biological properties endemic to the sunflower—has at least played a part in the turning of the sunflower (the attractive and regulative force of sunlight, of course, will have also played a part). Now let us consider Frankfurt's kindred example of a child moving a spider's legs by means of attached strings versus the spider *moving its own* legs.[97] In the first case, again, the mover of the spider's legs is obviously not the spider, but the child; in the latter case, the mover of the spider's legs clearly is the spider itself—that is to say, the dynamic complex of physiological conditions endemic to the spider. And so, on such a basis, we will discover among living organisms a pervasive (perhaps ubiquitous)[98] capacity for shallow, or minimally agential, behavior.

Yet there is also a general philosophical consensus—again beginning at least with Aristotle[99]—that human beings can exercise a more sophisticated or, as Korsgaard puts it, a "deeper" kind of agency. Korsgaard explains:

> [A]lthough there is a sense in which what a non-human animal does is up to her, the sense in which what you do is up to you is deeper. When you deliberately decide what sorts of effects you will bring about in the world, you are also deliberately deciding what sort of a cause you will be. And that means you are deciding who you are. So we are each faced with the task of constructing a peculiar, individual kind of identity—personal or practical

[97] Frankfurt (1988, p. 88).

[98] Insofar as all living things must display the key characteristics of life—the ability to grow, reproduce, metabolize energy, etc.—all such creatures would express an internal moving principle. See a list of these attributes here: https://www.reference.com/science/characteristics-living-things-d5fc0441ef59f417

[99] See Aristotle (2002, III.2).

> identity—that the other animals lack. It is this sort of identity that makes sense of our practice of holding people responsible, and of the kinds of personal relationships that depend on that practice. (2009, p. 19-20)[100]

There is a good deal packed into in this passage; I will limit my focus to what I regard as the two principal components of Korsgaard's conception of "deep" agency, that is, the "Second-Order" component and the related "Responsibility" component. First, recall that, with Kant, Korsgaard regards a non-human animal's behavior as "normatively loaded": that is to say, an animal's instincts or impulses *immediately* determine its behavior; the animal possesses no sufficiently well-developed self-conscious rational faculty, no transcendent, captainly "I", whereby it can *recognize* and *mediate* between its impulses and actions. [101] Persons, on the other hand, in virtue of their capacity to assume a self-conscious, first-personal, rational perspective, *can* mediate between impulse and action. Persons can *reflect* on their impulses and their motives and decide whether one is worthy of belief, or worthy of desires, on the basis of reasons. Indeed, recall that on Korsgaard's account, in order for a human being to act, she *must* act for reasons. This

[100] See also Korsgaard's further characterization of the "depth" of self-conscious human agency (2009, p. 19, my emphasis): "We are self-conscious in a particular way: we are conscious of the grounds on which we act, and therefore are *in control* of them. When you are aware that you are tempted, say, to do a certain action because you are experiencing a certain desire, you can step back from that connection and reflect on it. You can ask whether you should do that action because of that desire, or because of the features that make it desirable. And if you decide that you should not, then you can refrain. *This means that although there is a sense in which what a non-human does is up to her, the sense in which what you do is up to you is deeper*."

[101] As Kant puts it: "All animals have the capacity to use their powers according to choice. Yet this necessity is not free, but necessitated by incentives and *stimuli*. Their actions contain *bruta necessitas*." *Lectures on Ethics Collins*, (ed.) Peter Heath and J.B. Schneewind; (trans.) Peter Heath (Cambridge: 1997) 27:344, p. 125.

reflective activity constitutes the "second-order" component of deep agency. Which brings us to the "Responsibility" component. The idea is that, when I assume the role of a self-conscious "I" mediating among my facts, "I" am not, and cannot be, *identical* to any of the mental states or mental facts upon which "I" am reflecting or acting. As Velleman has put it, the role of the "single party" agent cannot be played "by anything that might undergo the process of critical review," because such a role "is precisely that it must be played by whatever *directs* that process" (2000, p. 139). Indeed, Velleman provides that the agent's "intervening between these items is not something that the items themselves can do" (2000, p. 125, emphasis mine). Now, insofar as "I," the "single party" play this role, "I" can author, or take possession of, my facts in such a way that the course and outcome of my deliberation will *not* constitute a mere expression, product, or outgrowth of my facts. Rather, the course and outcome will be "traceable directly to [me]".[102] In doing so, as Korsgaard puts it, I can "[insert] myself into the causal order".[103] That is, "I" *qua* agent, through rational deliberation, can determine the causal role that I *qua* empirical human being play in the natural, causal order. Thus, per the Rationalist account, we may understand an instance of human behavior as an expression of full-fledged[104] "deep" agency if it involves second-order reflection

[102] Velleman: "[W]hat makes us agents rather than mere subjects of behavior […] is our perceived capacity to interpose ourselves into the course of events in such a way that the behavioral outcome is traceable directly to us." David Velleman, *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000), p. 128, emphasis mine.

[103] Korsgaard puts the point with characteristic boldness: "The ideal of agency is the ideal of inserting yourself into the causal order, in such a way as to make a genuine difference in the world" (2009, p. 87)

[104] I have added the qualification "full-fledged" because different rationalists may possess different criteria for what constitutes deep agency. All of them will insist

and if one has self-consciously "authored" or "owned" her own attitudes or actions in such a way that her active causal contribution can be attributed *not* to forces at work *in* her or *on* her—but to her, "herself".[105]

Clearly, a good deal more will need to be spelled out here. Namely (1) what is involved in second-order "authorings" or "endorsements" of attitudes or actions; and (2) on what basis and to what extent might such authorship or endorsement render an agent responsible—that is, responsible in a way that sunflowers and spiders cannot be said to be responsible—for their behavior. To elucidate these ideas I will focus on two sorts of deep agency, what I will call "Productive" agency and "Complicit" agency. Accounts of *productive* agency will involve the actor *authoring*, *producing,* or *originating* (I will regard these terms as synonymous) an attitude or action. Accounts of *complicit agency* will involve an actor *assuming ownership* of *existing* attitudes or actions through her reflective endorsement of, or commitment to, them. Both sorts of deep agency will involve the "I" attempting, rationally and unilaterally, to intermediate among the facts of the "me" in such a way that, per the Rationalist account, will suffice for responsibility. Of course, I will challenge such accounts. That is, *critically*, I will *not* doubt that one can assume a second-order, agential stance with respect to

---

that second-order reflection is a necessary condition; all may not insist on responsibility being a necessary condition. Indeed, the conditional version of rationalism I will support here will advocate for the former but not the latter.

[105] Questions involving "agency" inevitably bleed into questions involving "free will," and the question of what metaphysical conditions must be satisfied such that an agent acts "freely." These investigations generally invoke two possibilities (1) Whether alternative possibilities are authentically open to an agent—otherwise known as PAP (the Principle of Alternative Possibilities), and (2) Whether the agent can be understood as the "source" of her own behavior. Here I will focus principally on "source" condition because it is more relevant to my thesis.

one's facts. This, I will take for granted. Rather, I will contest that, when one is engaged in self-reflection, the "I" should *not* be understood as *autonomously* and *unilaterally* acting on one's facts; rather the "I"'s activity should be understood as an *expression* of such facts or mental states. As such, I will argue against the responsibility condition.

4.3 Productive Agency

To more fully flesh out the nature of productive agency, let us turn to some characterizations in the literature. We will begin with Kant, who tells us:

> [E]ven if one believes the action to be determined by [natural]causes, one nonetheless blames the agent, and not on account of his unhappy natural temper, nor on account of the circumstances influencing him, not even on account of the life he has led previously; for one presupposes that it can be entirely set aside how that life was constituted, and that the series of conditions that transpired might not have been, but rather that this deed could be regarded as entirely unconditioned in regard to the previous state, as though with that act the agent had started a series of consequences entirely from himself. (1998: A555/B583)

Kant is describing a very natural tendency: When we blame a person for his action, we tend *not* to regard his action as a direct product or outgrowth of his facts or "the circumstances influencing him"—e.g., his history, his upbringing, his temperament, and the particular conditions bearing on him at the particular moment he performed his action. Rather, we tend to attribute the action *directly* to *him*—to the unitary self-reflective agent—, as if he, *qua agent*, has "started a series of consequences entirely from himself" (or "inserted himself into the causal order") and has done so in a way that is "efficient independent of alien causes." Recall, further, Sartre's assertion that "[i]t is I, always I, according to the ends by

which I illuminate these past events. Thus all my past is there pressing, urgent, imperious, but its meaning and the orders which it gives me I choose by the very project of my end." Here again, the suggestion is that the "I" can to some extent determine its ends *independent* of the factual history of the "me". And of course we find sympathetic accounts in the contemporary literature. Richard Moran, for example, asserts: "When we speak of 'authority' in connection with first-person statements of belief and other attitudes […] it is not just the report that the person is author of, but also, in a central range of cases, the person can be seen as the *author* of the state of mind itself, in the sense of being the person who *originates* it and is *responsible* for it" (2002, p. 113, emphasis mine). And Korsgaard affirms that while our practical identities are to some degree given to us by our cultures, societies, role structure, accidents of birth and natural abilities, we also "enter into them. And this means that desires and impulses associated with them do not just *arise* in us." In fact, she continues, "The motives and desires that spring from our contingent practical identities are […] *in part the result of our own activity*" (1996b, 239-240, emphasis added). Thus, according to the Rationalist account, at least in some cases, the authoring of a particular attitude should not be understood as an *outgrowth* or *expression* of factual circumstances at work *in* or *on* a person, but, rather, as the conjuring, or production, of the *person herself*.

But what kind of "hand" can we have in our exercise of agency, such that the production of a particular attitude can be attributed to "us" and not just to facts or circumstances at work "in" us or "on" us? For some helpful insight into the question, let us turn to these passages by Korsgaard and Moran.

> Korsgaard: [T]here is surely a difference between a case in which the event most immediately determining your movements is, say, that you are pushed from behind, and a case in which the event most immediately determining your movements, is a thought of your own. To take the most obvious case: most people do not feel that their freedom or power of self-determination is threatened by the possibility that their movements are determined by their own thoughts about what they ought to do. Rather, they feel that their freedom or power of self-determination is threatened by the possibility that this may *not* be the case. So perhaps we should claim that we are active to the extent that our movements are caused by our conceptions of what we ought to do. (2008, p. 11)
>
> Moran: My beliefs don't just happen to me; rather I am responsible for the reasons which I take to support them. This is part of what makes them mine.[106]

Notice that Korsgaard draws on the familiar distinction between something that pushes us—say a pair of human hands, a gust of wind, or even a strong impulse of fear or desire—and our own thoughts or "conceptions". These former "pushes," Korsgaard suggests, threaten our agency because they are *external* to us, or are, as Kant puts it, "heteronymous". Yet Korsgaard plausibly asserts that we do not regard our *own thinking* as external or heteronymous. After all, our thoughts do not seem to "impose themselves" on us or "just happen" to us: *we think them*. As Descartes has famously put it, "[i]t does not seem to me a fiction, but a truth which nobody should deny, that there is nothing entirely in our power except our thoughts; at least if you take the word 'thought' as I do, to cover all operations of the soul… not only meditations but acts of will."[107] As such, according to the Rationalist account, it is in virtue of our *active thinking* that we *take control* of the

[106] See *Contours of Agency* (eds.) Sarah Buss and Lee Overton, (Cambridge: MIT Press), 2002, p. 195.

[107] See Descartes' letter to Reneri for Polot (April 1638), from *Descartes: Philosophical Letters*, trans. Anthony Kenny (Oxford: Clarendon Press, 1970), p. 51. Passage found in Matthews (1992, p. 103).

raw material of our facts—our impulses, our desires and fears, and beliefs—and *determine* whether, for example, a particular impulse is worthy of being taken as a reason for a desire or a belief. Indeed, it is by means of such thinking, as Moran puts it, that an agent can "[orient] himself toward the question of his beliefs by reflection on what's true, or [orient] himself toward the question of his desires by reflecting on what's worthwhile or diverting or satisfying" (2002, p. 64).

Now, I grant that much of this account will square with our intuitions. It certainly *seems* to us that it is in virtue of our thinking that we "take command" of our attitudes and actions. And we do so, as the above-mentioned authors suggest, by directing our thought on our attitudes and determining whether they are worth following up on.[108] Yet, clearly the merit of this proposal turns on a crucial question, namely, whether we are indeed "orienting ourselves" to the facts of the case, or whether, indeed, the facts of the case are "orienting us". Note that this question is of critical importance, and this for the following reason: If the latter is the case, that is, if the facts of the case are orienting us, then our deliberative proceedings should *not* be understood as a product of the "I," but rather as a consequence of the facts of the "me".

[108] See O'Connor: "It does not seem to me (at least ordinarily) that I am caused to act by the reasons which favor doing so; it seems to be the case, rather, that I produce my own decisions, in view of those reasons." Timothy O'Connor, *Agent, Causes, Events* (ed.) T. O'Connor (New York: Oxford University Press), 1995, p. 196.

4.3.1 The Case of John's Career

Let us thus turn to the case of *John's Career* and consider a number of proposals whereby John will ostensibly "orient himself" to the facts of the case and thereby "produce" a propositional state—in this case, a belief about what career will be in his best interest to pursue.

John has just finished his undergraduate degree and is trying to decide which career to pursue. At this time, to employ Korsgaard's conceptual framework, John associates himself with Practical Identity 1: I'm a person who honors my parents' wishes and who therefore endorses the following Principle of Choice (a): "Honor your parents' wishes." And suppose, following up on this principle, John, while not particularly interested in a medical career, has nonetheless honored his parents' wishes and applied to, and been accepted by, a prestigious medical school. In the meantime, an admissions scout from an MFA program has discovered John's paintings hanging on a coffee shop wall and has offered John a full scholarship. As a matter of fact, John also associates himself with Practical Identity 2: I'm part of a group of artistic bohemians who abides by Principle of Choice (b): "Follow your passion." And John knows he will probably derive much more personal satisfaction from a career in the arts. Finally, John also identifies with Practical Identity 3: I'm a grown man and I am my own person, and, as such, I can and should make up my own mind on the basis of my own reasons. As such, John 3 respects Principle of Choice (c): "Never be swayed by mere impulses. Always make up your own mind, and do so on the basis of your own reasons." Now, on Korsgaard's account, for John—and not merely the facts

at work within John—to determine which future path he should take, John, operating as John (3), will need to take a step back, review both career paths and then arrive at a principled decision as to which choice he should ultimately endorse. But how exactly will John go about doing this?

Suppose John first conducts a cost-benefit analysis. John composes a list of considerations that weigh in favor of and against each candidate career choice. Having composed the list, John reviews both sets of considerations in an effort to ascertain their merits. Suppose John encounters the consideration: "If I don't earn the medical degree, I will have rejected my parents' advice and will thereby disappoint them, not to mention wasted all the money they invested in my undergraduate program." This consideration strikes John as a very good reason to pursue the medical career. In fact, the moment the prospect of disappointing his parents occurs to John he is overcome with a wave of dread and feels a powerful impulse to pursue the medical career, however uninspiring. Yet at that very moment the voice of John's Practical Identity (2) kicks in: "Hold on! What about your creative passion? Do you feel no obligation to follow your artistic calling? Imagine how empty, how pitiful you will feel if you follow the conventional straight and narrow, just for fear of disappointing your parents!" Again, this thought strikes John as eminently reasonable. Moreover, the prospect of living in such a "compromised" fashion induces another wave of dread and generates an impulse to reject the medical profession. Yet now the voice of John (3) strongly asserts itself: "Oh, c'mon! You're getting all bent out of shape here, bullied about by your emotions. Just calm down, pull yourself together and take *a cool,*

impartial look at each career path and corresponding principle of action. Just look at the facts and arrive at an objective conclusion." And so John (3) attempts to do just this. He takes a deep breath, rubs his eyes, and asks himself, afresh, which is *really* the most reasonable: (1) Honoring his parents vs. (2) Following his passions. He may even ask himself, as Korsgaard (via Kant) suggests, which principle would best apply, in general, to *any* college graduate in his position—such that he can will one principle or another to operate as a "law". Yet the moment John (3) attempts to arrive at such a verdict, he is stampeded by the same mob of anxieties that plagued him before, and he finally collapses in a heap of futility and despair.

I hope that this portrait of John's unsuccessful efforts at practical reasoning will strike a familiar chord. After all, deliberation—especially concerning complex questions of great importance—can be extremely difficult, and we rarely meet with as much success as we wish. Nonetheless, note that John is clearly and actively *doing* all sorts of things. He has *focused* on the career question before him; he has *written down* the list of considerations as they occurred to him; and he has tried impartially to ascertain the merits of each consideration and the merits of each principle, in general. All of these activities might properly be understood as John's efforts to orient himself to the facts of his case, or, as Moran put it, to "orient himself to his reasons." And yet, to what extent has John truly "interposed" himself[109] in his mental processes? That is to say, to what extent has

[109] Recall Velleman's passage: "[W]hat makes us agents rather than mere subjects of behavior […] is our perceived capacity to *interpose ourselves* into the course of

John, operating as John (3) *intervened* among his facts in a way such that his activity should be attributed "him" and *not* to his facts?

Let us take a closer look. Notice, first, that John's idea to write up a cost-benefit analysis just *occurred* to him. As a matter of fact, when John needs to make up his mind about difficult matters, usually does any number of things. Sometimes he calls a friend, or goes for a walk, or writes in his journal, or takes a nap, or composes a cost-benefit analysis. In this case, the appeal of composing a cost-benefit analysis just spontaneously *occurred* to him and won him over. Note, secondly, that as John composed his lists of considerations, these considerations just *rose up* in his mind; he "asked himself" what were the costs and benefits of each proposal, just as one may pose a question to a Magic Eight Ball, and the responses just rose up to the "window" of his attention. Further, in focusing on each consideration and attempting to evaluate their merits, or their "reasonableness," note that the course of John's cogitations, or "mental ballistics" as Strawson puts it,[110] proceeded rapidly and out of view and were likely influenced by any number of preferences, fears and desires, of which John is not conscious. As such, the deliverances of such cogitations, again, just occurred to him, as did their attendant emotions.[111] Moreover, note that John did not

---

events in such a way that the behavioral outcome is traceable directly to us" (*The Possibility of Practical Reason*, 2000, p.128, emphasis mine)

[110] See Galen Strawson, "Mental Ballistics and Second Order Belief," *Proceedings of the Aristotelian Society N.S.*, (2003) 103: 227-56.

[111] Nietzsche is especially good here. See (2001, p. 112, §111): "The course of logical thoughts and inferences in our brains today corresponds to a process and battle of drives that taken separately are all very illogical and unjust; we usually experience only the outcome of the battle: that is how quickly and covertly this ancient mechanism runs its course in us."

voluntarily invoke his practical identities; they arose as if in response to the prompting of occurring fears and desires. For example, John's recognition of what he regards as the sell-out tedium of a medical career excited his "retaliatory" longing to pursue the passions of the artistic career (Practical Identity 2), which, in turn, prompted his fear of disappointing his parents and brought back into play Practical Identity 1. Indeed, even the voice of Practical Identity (3) seemed to wax and wane in relation to the other voices. As such, while John did perform various activities in an effort to "orient himself to reasons" or to "author his own reasons," very little of this activity proceeded according to the direction a distinct, self-conscious authoritative "I"; rather, it proceeded spontaneously, "beneath the hood," so to speak, of John's conscious mediation.

Now, suppose John begins to suspect this. That is, suppose John begins to sense that, even when he has attempted to "pull back" and don the mantle of the supervisory "I," this "I", as Arpaly has puts it, has been acting as a "dupe" of "other desire*s*, emotions, or irrationality". Or, he suspects, as Arpaly puts it, that "the mere first person *experience* of [his] having control over [his] mental life is not by itself a surefire indication that [he] *actually* has control over [he] mental life, in any meaningful sense of 'self-control'" (Arpaly, 2003: 19). And suppose this revelation induces a fresh wave of panic. After all, like many of us who desire to become "bounded, masterful, autonomous selves," John wishes that he "himself"—that is, John operating as John 3—should serve as the responsible chaperone of his mental activities. And so, in fit of frustration, John resolves to *really bear down* and do *whatever is required* of him to assume firmer hold of the

deliberative reins, making absolutely certain that it is *he himself* and not *forces within him* that will determine the course and outcome of his deliberations.

How will John go about doing this? Let us explore two natural-seeming proposals: "Scratch" and "Decision."

4.3.2 Scratch

In the first case, to make sure "he himself" and not just "forces within him" has determined the course and outcome of his deliberation, let's suppose that John resolves to take a much larger step back. That is, suppose John recognizes that his previous effort to occupy the perspective of Practical Identity (3) was a failure. After all, recall that when he attempted rationally and impartially to resolve the tensions between Practical identities (1) and (2), John succumbed to a tempest of emotions. So this time John resolves to summon all his strength and *evacuate*, as much as possible, the emotion-laden Practical Identities (1) and (2) and fully inhabit Practical Identity (3) such that he can engage his "pure reason" which, as Kant has suggested, can "of itself, independently of anything empirical, determine the will" (1997b, 5:42). From such an untarnished, luminous, vantage point, from such a "neutral substratum,"[112] as Nietzsche puts it, John will make certain that his

[112] See Nietzsche's criticism of such a "neutral substratum" in discussion of lambs and birds of prey (2006, p. 26, §13). "For just as the popular mind separates the lightning from its flash and takes the latter for an action, for the operation of a subject called lightning, so popular morality also separates strength from expressions of strength, as if there were a neutral substratum behind the strong man, which was free to express strength or not to do so. But there is no such substratum…"

deliberations proceed free of any subversive inputs issuing from the other two, seditious, practical identities.

Yet, as discussed in the previous chapter, we should rightly be skeptical of John's prospects. Recall the formidable impediments that plague any effort to achieve an accurate grasp of one's facts. Recall the various experiments in which subjects were informed of common epistemic blind spots and were asked to review their reflections to make certain that they themselves had not succumbed to them—only to deny their susceptibility to such blind spots and arrive at the very same epistemically compromised conclusions. As such, in spite of John's most conscientious efforts to overcome the influence of biases stemming from his "intellectual character," or his "contingent profile of concerns," as Blackburn puts it[113] his analysis will still, in all likelihood, yield to a number of undetected (and undetectable) prejudices. Furthermore, as previously discussed, John's attempts to establish or infer new facts about himself—whether, for example, he *really should* prefer one principle over another—will also likely succumb to such influences.[114]

Yet concerns loom from other directions. Consider the following normative worry. First, recall that on Korsgaard's view, a person *must* act for a reason; indeed, as Korsgaard has said, one's reasons must go "all the way down" (2003, p.118). But by what means will John discover reasons that will *ultimately* justify one practical identity or principle of choice over another? That is, how, by appeal

[113] See Blackburn (1998, pp. 241)

[114] Again see Nietzsche: "[M]ost of a philosopher's conscious thought is secretly directed and forced into determinate channels by the instincts. Even behind all logic and its autocratic posturings stand valuations, or stated more clearly, physiological requirements for the preservation of a particular type of life." (2002, p. 7, §3) and also (2002, "On The Prejudices of Philosophers," p. 5-24)

to "reason alone," might John impartially weigh the merits of his two competing principles, (1) Honoring one's parents, and (2) Following one's passion? Surely, reason will permit John to amass a great deal of data; it will permit him, as Hume suggests, to perform various inferences, mathematical calculations and exercises of demonstrative reasoning;[115] it will allow him to ascertain causes and effects associated with various proposals and "[regard] the abstract relations of [his] ideas" (*T*: 2.3.3, p. 265). It will permit him to calculate various probabilities, and make various projections. But, having amassed all such data, where will John discover a final, impartial reason that will establish, all things considered, the *preferability* of one principle over the other? Note, as Blackburn says, that for a reason to derive from Reason alone, it must be "capable of appealing to all reasonable people simply in virtue of their rationality, and independent of any particular desire or interest they happened to have" (Blackburn, 1998, p.253). And yet, as Nagel has pointed out, eventually the question of why one set of considerations will win out over another set will eventually "have no answer or it will have an answer that takes us outside of the domain of subjective normative reason and into the domain of formative causes of [one's] character" (1986, p. 117). That is to say, since nothing "impartially" recommends one principle over another,[116] John eventually will have to tap into his "contingent profile of

[115] David Hume, *A Treatise of Human Nature*, (ed.), David Fate Norton and Mary J. Norton, (Oxford: Oxford University Press, 2005), (2.3.3, p. 265). Henceforth: "*T*".

[116] At least, it is very difficult to understand how any reason could possibly, impartially, determine whether it is, in general, preferable for a person to honor her parents or to pursue her passion. As Hume has famously remarked: "'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of

concerns," or, as Williams puts it, his "motivational set".[117] As such, John's preference will eventually "bottom out" in the question of how much he loves his parents and fears disappointing them, or how passionately he feels about a career in the arts and fears a less-than-passionate life. Yet note that such fears or desires themselves are likely not the product of conscious reflection. That is, John's desire for his parents' approval is not something John has "authored" or "produced" as a result of practical deliberation; nor is the pleasure he takes in his artistic work or his asphyxiating fear of abandoning such work. Rather, such preferences have accrued to him by way of nature and upbringing, and are as un-attributable to his authoritative self-conscious deliberative will as his distaste for olives or his fear of the dark. In fact, over the course of John's life, John has unselfconsciously built up a vast number of dispositions[118]—longings, fears, prejudices—irrespective of self-conscious rational scrutiny and endorsement. As such, to the extent that John's reflection is influenced by his un-authored

my finger." (*T*, 2.3.3, p. 267). And Blackburn has put it, "There is no *necessary* object of concern" (1998, p. 253).

[117] (Williams, 1981) Frankfurt has also put this thought nicely. Speaking of the Kantian "pure" will, he writes: "This pure will is a very peculiar and unlikely place in which to locate an indispensable condition of individual autonomy. After all, its purity consists precisely in the fact that it is wholly untouched by an of the contingent personal features that make people distinctive, and that characterize their specific identities […] The pure will has no individuality whatsoever." (Frankfurt, 1999, p. 132). See also Williams: "[P]ractical deliberation is first-personal, radically so, and involves an *I* that must be more intimately the *I* of my desires." (Williams, 1985, p. 67)

[118] See Michael Smith's illuminating treatment of his "dispositional conception of desire". Michael Smith, "The Humean Theory of Motivation," *Mind*, New Series, Vol. 96, No. 381 (Jan., 1987) pp. 36-61, esp. pp. 50-54.

preferences, his practical deliberation cannot properly be understood as a pure product of rational autonomy.[119]

Indeed, the influence of un-chosen dispositions will give rise to a further, familiar, metaphysical problem, what I will call the "Problem of Self-Creation". Nagel has put the problem thus:

> By increasing our objectivity and self-awareness, we seem to acquire increased control over what will influence our actions, and thus to take our lives into our own hands. Yet the logical goal of these ambitions is incoherent, for to be really free we would have to act from a standpoint completely outside ourselves, choosing everything about ourselves, including all our principles of choice—creating ourselves from nothing, so to speak.
>
> This is contradictory: in order to do anything we must already be something. (1988, p.118)[120]

That is to say, as discussed above, insofar as John wishes to exercise rational autonomy over his deliberations and, in so doing, produce or author his resultant attitudes *whole cloth*, he will also need to author those preferences that feed into and influence his deliberations (for, otherwise, un-authored preferences will influence his process). Yet the prospect of "getting behind" one's entire ensemble of existing preferences is problematic, for, from such a neutral position, from such

---

[119] See Arpaly's rather extreme view: "[E]very step I take in deliberation is informed in a non-deliberative way by beliefs and desires that do not participate in it" (2003: 59).

[120] See Blackburn's and Williams' versions of this "Self-Creation" argument. First Blackburn: "You, when you deliberate, are whatever you are: a person of tangled desires, conflicting attitudes to your parents, inchoate ambitions, preferences and ideals, with an inherited ragbag of attitudes to different actions, situations, and characters. You do not manage, ever, to stand apart from all that" (1998, p. 252). Williams: "The I that stands back in rational reflection from my desires is still the I that has those desires and will, empirically and concretely act; and it is not, simply by standing back in reflection, converted into a being whose fundamental interest lies in the harmony of all interests. It cannot, just by taking this step, acquire the motivation of justice" (1985, p. 69).

a "view from nowhere,"[121] one would possess no preferential basis according to which his deliberative process could even get off the ground. Thus, as Nagel puts it: "In order to do anything we must already be something." Or, as Frankfurt has put it, once we occupy such a preference-shorn standpoint, there is "no fixed point from which a self-directed volitional process can begin" (1999 p. 110). As such, Frankfurt asks: "What preferences and priorities are to guide him in choosing, when his own preferences and priorities are among the very things he must choose?" And he answers: "It appears that he is left with so little volitional substance that no choice he makes can be regarded as originating in a nature that is genuinely his. With respect to a person whose will has no fixed determinate character, it seems that the notion of autonomy or of self-direction cannot find a grip" (1988, 177-178). [122] Thus we find that the "I," in an effort to cut loose its empirical strings, eventually discovers it has lost its volitional basis: in realizing its freedom, it has immobilized itself.

4.3.3 Brute Decision

It thus appears that the effort to "evacuate" John 1 and John 2 and fully occupy the purely rational John 3 in an effort to render a decision will, at the very least, encounter formidable resistance. But perhaps John can circumvent these

[121] Nagel (1986). See, also Galen Strawson's concise treatment of this in his articulation of his "Basic Argument". Galen Strawson, "The Impossibility of Moral Responsibility," *Philosophical Studies*: *An International Journal for Philosophy in the Analytic Tradition*, Vol. 75, No. ½ (Aug. 1994) pp. 5-24.

[122] See also Williams: "I am, at the time of mature reflection, what I have become, and my reflection, even if it is about my dispositions, must at the same time be expressive of them. I think about ethical goods *from* an ethical point of view that I have already acquired and that is part of what I am." (1985, p. 51)

concerns and determine his attitudes via another route. That is, perhaps John can just override the arduous process of deliberation and simply, spontaneously, make it the case, *just like that*, that he believes one career to be preferable, or else make it the case, *just like that*, that he desires one career over the other, by means of "brute volition".

At least when it comes to the determination of beliefs, Ginet[123] has suggested that we commonly do this very thing. Consider the following case[124]. You have headed out on a road trip, and an hour into the trip, you wonder if you have left the oven on. You know it is indeed possible that you have left it on (though you rarely do leave it on, you have in fact done so enough times to legitimize your worry), but it is also likely that you have in fact turned it off. It is too late to turn back and you know of no one who can check for you. And so, for the sake of psychological expedience, you "just decide" to believe that you have turned the oven off. In this case, you will *render it the case* that you possess a certain attitude—your belief—purely by dint of will, that is, by means of a voluntary decision. In a like manner, John might "just decide," by dint of will, to believe that a particular career path is preferable.

Yet this proposal has prompted a number of objections. Alston, for example, suggests that in such cases one is in fact subconsciously responding to evidence that favors one prospect over the other.[125] Moreover, on Alston's view, it is

[123] Carl Ginet, "Deciding to Believe." In *Knowledge, Truth, and Duty*, (Ed.) Matthias Steup, (Oxford: Oxford University Press, 2001) p. 63-76.
[124] This case is based on Ginet's example discussed in 2001.
[125] For a similar argument, see Nottleman (2006, p. 572-573). For an argument that Ginet's agent is only "accepting" and not "believing" a proposition, see

simply a contingent fact of our human, cognitive wiring that we cannot "just decide" to believe a proposition. As he puts it: "When I look out my window and see rain falling, water dripping off the leaves of trees, and cars passing by, I no more have immediate control over whether I accept those propositions than I have basic control. I form the beliefs that rain is falling, etc. willy-nilly."[126] As such, Alston contends, we are just wired in such a way that we cannot exercise deliberate control over our beliefs.

Others[127] have made a stronger case, arguing that the voluntary, decisive formation of a belief is a *conceptual* impossibility. Let us consider one such argument developed by Patricia Hieronymi. Hieronymi claims that "you cannot, properly speaking, form and execute an intention *to believe*. You can, at best, form and execute an intention *to bring it about* that you believe" (2009, p.157). To make her case, Hieronymi first distinguishes beliefs from *supposings* or *imaginings* in virtue of the fact that "to believe that p is to be committed to p as

---

Andrei A. Buckareff, "Acceptance and Deciding to Believe," *Journal of Philosophical Research*, Volume, 29, 2004.

[126] William Alston, "The Deontological Conception of Epistemic Justification." In *Essays in the Theory of Knowledge* (Ithaca: Cornell University Press, 1988), p. 115-152. Alston distinguishes "basic" from "immediate" voluntary control as follows. Basic control is a type of control we exercise when we when we simply decide to perform an action, and our action follows immediately upon our decision. He suggests that we can exercise immediate non-basic control over beliefs that arise as a result of "one uninterrupted intentional act" (that is, beliefs which arise, not merely as a result of a decision, but as a result of one or a series of actions).

[127] For a good sampling, see: Bernard Williams, "Deciding to Believe" from *Problems of the Self* (Cambridge: Cambridge University Press, 1974) p.136-151; Dion Scott-Kakures "On Belief and Captivity of the Will." *Philosophy and Phenomenological Research* 54 (1994): 77-103 and Pamela Hieronymi, "Controlling Attitudes." *Pacific Philosophical Quarterly* 87 (2006), p. 45-74. See also Hieronymi "Believing at Will," from *Belief and Agency*, David Hunter, ed., *The Canadian Journal of Philosophy* Supplementary Volume 35 (2009), p. 153.

true—to take p to be true in a way that leaves one answerable to certain questions and criticisms […] to standards of justification, warrant, or consistency that govern belief" (2006, p. 49).[128] Further, Hieronymi distinguishes between two kinds of "reasons for belief"—reasons that bear directly on the truth of p, or, as she puts it, *constitutive reasons*, and reasons that do not vouch for the truth of p but may suggest that p is a *beneficial* belief to hold, what she calls *extrinsic reasons* (pragmatic reasons to believe p would thus qualify as extrinsic—as in the Ginet case mentioned above). Hieronymi explains that when one "intends to believe p" one must obviously *not already believe* p, for if one already believed a proposition, one could not simultaneously "intend" to do so. Rather, on Hieronymi's account, when one "intends to believe p" one desires to believe p on the basis of extrinsic reasons—that is, on the basis of reasons that argue for the desirability, not the truth, of p. And yet, one "cannot become committed to an answer to a question by finding convincing reasons that you, yourself do not take to settle *that* question" (2009, p. 165). That is to say, it appears to be a conceptual truth that if I am seeking an answer to a question of whether p, and if I encounter reasons or evidence that I do not take to directly address the question of whether p, I cannot take such reasons or evidence *to answer* the question. For this reason, Hieronymi claims that, "you will not, by finding convincing reasons that you take only to show that believing p is worth doing, therein become committed to the

[128] On the question of the "truth-directedness" of beliefs, see, for example, Nishi Shah and David Velleman, "Doxastic Deliberation." *The Philosophical Review*, (Oct. 2005), Vol.114, No. 4. See also David Velleman "On the Aim of Belief," from *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000) and Nishi Shah "How Truth Governs Belief" *The Philosophical Review*, (October 2003)Vol. 112, No. 4.

truth of p" (2009, p. 165). Indeed, one might in fact commit oneself to the proposition that p is worth believing, and subsequently *cause oneself* to believe p, but, in such a case, Hieronymi asserts that the eventual acquisition of the belief in p (that is, the commitment to *the truth* of p) should be understood "not as a part of the action you have decided upon, but rather as the product or consequence of that action—an action best described as bringing it about that you believe or making yourself believe" (2009, p.163).

While these may not be knockdown arguments, I suggest that, both for contingent and conceptual reasons, the proposal that one can "just believe" or just "make oneself believe," a proposition, does not seem promising. Rather, the extent to which we believe a proposition seems constrained by our appreciation of the truth. Will we fare any better when it comes to "making ourselves desire" just like that? That is, can we render it the case that we desire x? Unfortunately, this seems even less plausible. After all, it just seems obvious that I cannot decide that I prefer anchovies to mushrooms on my pizza, or that I prefer Britney Spears to Beethoven. Or, indeed, I may certainly *decide* to render it the case that I desire such things, but this decision won't, there and then (or, in the latter case, hopefully ever), render it the case that I *actually do* possess such desires.

Perhaps Velleman can be of some assistance here. In his essay, "What Happens When Someone Acts," Velleman suggests that among an agent's most fundamental desires is the desire to act in a way that appears the most reasonable. So, if an agent comes to realize that an initially *less* desirable option A ultimately is more reasonable than initially *more* desirable option B, the agent can "throw his

weight behind the motives that provide the strongest reasons,"[129] in favor of option A, and thus render his weaker desire the stronger. "For when a desire appears to provide the strongest reason for acting," Velleman explains, "then the desire to act in accordance with reasons becomes a motive to act on that desire, and the desire's motivational influence is consequently reinforced" (1992, p. 141). So, suppose Greg is on a gluten-free diet and a friend presents him with the option of having a slice of birthday cake or an apple. The cake may initially appear more desirable, but, after reflecting on his health, and the consequences of eating gluten, and his resolution to stop eating it, set against the momentary pleasant satisfaction of eating the cake, Greg concludes that the prospect of eating the apple is more reasonable than that of eating the cake. Since Greg desires to act for the best reasons, and since eating the apple presents itself as the most reasonable option, Greg "throws his weight" behind his minimal desire to eat the apple, and, in doing so, renders the apple the more desirable option.

Now, I will have significantly more to say about cases like this in Chapter 7. But, for now, let us note just a few things. To begin with, note that, per Velleman's account, Greg has *not* actually rendered it the case that option A is more desirable, *just like that*. Rather, what has rendered A more desirable has been Greg's *coming to the conclusion* that A is more reasonable and *this conclusion* connecting up in the appropriate way with a pre-existing desire—that is, a desire to act on the basis of the best reasons. Note, second, that it is just a *matter of luck* that Greg's desire to do the most reasonable thing is sufficiently strong such as to render the apple more desirable (after all, we may easily imagine

[129] Velleman (1992, p. 141)

a scenario in which Greg's desire to do the most reasonable thing yields to a *stronger* desire for the comfort and satisfaction of eating the cake). Third, note the ambiguity involved in Velleman's characterization of Greg "throwing his weight" behind his more reasonable desire to eat the apple. Again, I will have more to say about this in Chapter 7. But for now, I will suggest that what actually takes place is as follows. Greg's *desire* to do the most reasonable thing is activated by his conclusion that the most reasonable thing to do in this case is to eat the apple, and this desire *automatically* reinforces his additional desire to eat the apple.[130] As such, "Greg himself," does not "throw his weight" behind his desire to eat the apple; indeed, the "transfer of power" does not require the involvement of "Greg" "throwing his weight" at all. Finally, notice that, in concluding that eating the apple is the more reasonable thing to do, Greg has in fact *not* actually developed a stronger desire for the apple, *per se*. Rather, the object of Greg's desire has just been to "do the most reasonable thing" which happens to involve eating the apple—the apple *itself* may not strike him as any more desirable. Indeed, I can imagine a scenario where I conclude that the most reasonable thing to do would be to listen to Britney Spears music for two hours (suppose, for example, I need to take my kids clothes shopping at the mall), but, having arrived at such a conclusion, I will still find the pumped-in music deplorable. As such, I have argued that it is one's conclusion to do the reasonable thing—as opposed to any activity on the part of a distinct "agent"—that renders one capable of desiring p;

[130] Indeed, this seems to follow from Velleman's remark that "the desire's motivational influence is consequently reinforced." Such a "hydraulic" characterization seems to contradict his assertion that the agent himself has to do any "weight throwing".

or else that one's desire to do most rational thing does not render the object of desire "more desirable," *simpliciter*. In any case, in no sense do we find an autonomous agent rendering it the case that she decides to desire p *just like that*.

* * *

I have thus argued that "Scratch," or the endeavor to "get completely behind" oneself in order to author or produce an attitude, is enormously difficult, if not conceptually incoherent. I argued that, given that our doxastic mechanisms are either contingently or conceptually "wired" to our appreciation of truth, we cannot "just decide" to believe a certain proposition. And I have suggested that we are no more capable of "just deciding" to produce or enhance a desire, for our desires are, by and large, hardwired, and not susceptible of voluntary decision or manipulation. Indeed, even Velleman's formulation suggests that our "deciding to desire p" is contingent upon the collusion of our conclusion that p is "more reasonable," and the triggering of a sufficiently powerful pre-existing desire to act on one's "superior rational force".

Yet perhaps one does not need to "author" or "produce" an attitude in order to exercise responsible agency. Perhaps our "insertion into the causal order" can be located in a different place, that is, in our exercise of "complicit" agency, where one can "assume ownership" of one's attitudes and actions. So, let us now take a look at some such "ownership" accounts. I will focus, first, on the compatibilist accounts of Fischer and Ravizza and Lynne Baker. I will then revisit Moran's notion of avowals.

4.4 Complicit Agency: Fischer and Ravizza on Ownership

Fischer and Ravizza agree that a person's orientation to reasons is a necessary component of " deep" agency. But Fischer and Ravizza do not focus on a person's capacity to produce or author an attitude, a capacity for what they call "Regulative Control"[131]. Rather, on their account, an agent's responsibility-conferring self-conscious intervention requires only what they call "Guidance Control," a control that requires of one's reasons-response mechanism (i) That it be appropriately[132] responsive to reasons (1998, p. 89); and (ii) That it be "one's own" (1998, p. 99). How does a person come to "own" her reasons-response mechanism? Fischer and Ravizza provide that a person does so when she satisfies three conditions: (i) An individual must see herself as the source of her behavior; (ii) She must accept that she is a fair target of the reactive attitudes that result from how she exercises her agency in certain contexts; and (iii) Her view of herself must be based, in an appropriate way, on the evidence (1998, p. 210-214).

Now, I will concede straightaway that Fischer and Ravizza's view has much to recommend it. Indeed, for reasons I will more fully develop in chapters Seven and Eight, I believe their compatibilist account of responsibility may represent all that

---

[131] Strictly speaking, their notion of regulative control involves being able "to do otherwise," in the sense that an agent is metaphysically free to choose other than the way she actually chooses (something that determinism rules out). I suggest that regulative control might also be understood as a kind of Productive Agency, that is, the capacity to decide one's outcome irrespective of one's antecedent personal facts.

[132] Note that I will be making specific reference here to Fischer and Ravizza's characterization of "moderate reasons-responsiveness" as opposed to both "strong reasons-responsiveness," which they consider "too strong" a condition for moral responsibility (1998: p. 89) and "weak reasons-responsiveness" which they consider "too weak" a condition for moral responsibility (1998: p. 89).

we, time-bound, flesh and blood creatures can reasonably expect to achieve and exercise. After all, the ambition to take ourselves completely "in our own hands" or to "create ourselves from scratch" does seem unrealistic, if not incoherent. Yet we certainly can, and indeed, we *must*, as Fischer and Ravizza claim, "assume ownership" of, or "assume responsibility" for, our attitudes and our actions. Indeed, so long as we wish to participate in our social world with its customs of promise-keeping, law-abiding and forgiveness, we *must* do so.

In spite of the ostensive virtues of Fischer and Ravizza's account of responsibility, one may nonetheless find it unsatisfying or "shallow," as Smilansky puts it[133] (or, indeed, regard it as a form of "subterfuge," as Kant see it).[134] And this for a number of reasons. To begin with, again, Fischer and Ravizza's account does not provide for the "production" or "authoring" of one's behavior or attitude, and some[135] regard this as a necessary condition for *ultimate* blame and credit grounding responsibility. Further, the very assumption of ownership clearly does not link up in any sufficient way with genuine responsibility. As Nietzsche has famously put it: "We laugh at him who steps out of his room at the moment when the sun steps out of its room, and then says: 'I *will* that the sun shall rise'; and at him who cannot stop a wheel, and says: 'I *will* that it shall roll'; and at him who is thrown down in wrestling, and says: 'here I lie

[133] Such is Smilansky's (2003) criticism.
[134] Kant (1997, 5:96, p. 81)
[135] Such did Kant. Also see Strawson (1994), Pereboom (2001), Smilansky (2003).

but I *will* lie here!'"[136] Indeed, my "assuming ownership" of my special disorientation or my deformed feet clearly does not render me any more genuinely responsible for having developed such liabilities. Yet, of course Fischer and Ravizza's account does not focus on the rising of the sun, or on such specific bodily or psychological afflictions, but rather on one's "reasons-response mechanism". As such, any proper criticism will need to focus here. So focus here, I shall.

Note first their three conditions on assuming ownership of one's reasons-response mechanism, that (i) An individual must see herself as the source of her behavior; (ii) She must accept that she is a fair target of the reactive attitudes that result from how she exercises her agency in certain contexts; and (iii) Her view of herself must be based, in an appropriate way, on the evidence. Of course much will ride on condition (iii), that is, the question of whether an agent's view of herself has been based, in an appropriate way, on the evidence. For one might very well form a view of oneself that satisfies (i) and (ii), but that is insufficiently or incorrectly based on the evidence, or based on the wrong kind of evidence. As such, Fischer and Ravizza's account will require a clear specification of what it might mean for a person's view of herself to be "appropriately based" on the evidence. Unfortunately, Fischer and Ravizza do not offer much help here; rather, they remark that "[t]he specification of the relevant notion of appropriateness here

[136] Nietzsche, *Daybreak*, (eds). Maudemarie Clark and Brian Leiter, (trans.) Hollingdale, R.J., (Cambridge: Cambridge University Press, 1997), p. 77, §124.

is a delicate and difficult matter" (1998, p. 213), and that "the relevant notion of appropriateness must remain unanalyzed" (1998, p. 236).

Fischer and Ravizza's failure to provide such an adequate specification prompts a number of concerns, especially when it comes to manipulation cases. That is, in cases that involve hypnosis, or drug-induced states, or the work of a mad neuroscientist, Fischer and Ravizza concede that an agent *cannot* properly recognize the effect of such manipulations, and hence, *cannot* responsibly "own" her reasons-response mechanism. Thus they concede that someone who, for example, has been "electronically induced to have the relevant view of himself […] has not formed his view of himself in the appropriate way" (1998, p. 236).[137] To see why this presents a problem, take the case of *Hypnotized Bill*. Bill has been hypnotized to believe (1) That he is a re-incarnation of Napoleon and, as Napoleon, that he is the source of his own behavior; (2) That people will accuse him of foolishness and will tell him that he has been hypnotized into believing that he is Napoleon, and they will do this because they are afraid and envious of his supreme authority; and (3) That, as a result, such teasing is sure-fire evidence of the fact that he really is Napoleon and that his conviction is *not* the product of hypnosis. Under such conditions, Bill/Napoleon will indeed regard himself, as he marches through town, as the "source of his behavior," and as "a fair target of the reactive attitudes that result from how he exercises his agency in certain contexts." And yet, insofar as Bill's beliefs have been implanted in him, it is clear that Bill's view of himself has *not* been based appropriately on the evidence and,

[137] Fischer and Ravizza assert that one who is "electronically induced to have the relevant view of himself […] has not formed his view of himself in the appropriate way" (1998, p. 236).

thus, that Bill is in no position to genuinely assume ownership of his reasons-response mechanism. Indeed, as noted, Fischer and Ravizza concede this point.

But in conceding this point, Fischer and Ravizza expose the chief vulnerability of their position. That is, if we allow that such a case of manipulation undermines one's ownership of one's reasons-response mechanisms, one may wonder whether the cumulative effect of enculturation, habituation, or peer pressure, may pose a similar threat. For recall that in manipulation cases what principally threatens the status of one's "ownership" is precisely the fact that, due to the manipulated or "implanted" nature of one's reasons-response mechanism, one cannot properly claim ownership of one's reasons-response mechanism. That is, in such a case, one cannot responsibly assume ownership of one's reasons-response mechanism because to do so one would involve the implementation one's very reasons-response mechanism, the corruption of which, again—due to the nature of its corruption—one cannot detect.[138] Indeed, as we discovered in our examination of the pitfalls of self-knowledge, we are very commonly led to conclusions by means of pressures, suggestions or unconscious mechanisms that we have not, and, in many cases *cannot* recognize.[139] Of course there is no shortage of cases where persons young and old have been raised in a closed environment and have been

[138] Recall Blackburn: "We can compare the situation to looking at our own eyes in a mirror. We might see that our eyes are cloudy; but if they are, it will be with cloudy eyes that we see it" (1998, p. 261). And see Wittgenstein, speaking of a similar problem: It is "[a]s if someone were to buy several copies of the morning newspaper to assure himself that what it said was true." Wittgenstein, *Philosophical Investigations*, (trans.) G.E.V. Anscombe (Blackwell Publishing Ltd.), §256.

[139] Derk Pereboom nicely develops this case in his "Four Case Argument". See Derk Pereboom (2001: pp. 112-126)

trained to embrace certain propositions, and, in addition, to believe that they themselves have responsibly arrived at their own feelings and conclusions. As a particularly vivid example, consider the case of "The Most Hated Family in America" [140]—the Phelps family, who are the principle members of the Westboro Baptist Church. The Phelps children are brought up to believe that, for example, the United States is an immoral country due to its tolerance for homosexuality. As such, family members gather and picket military funerals, blazoning signs that deprecate American soldiers and homosexuals. Moreover, the Phelps' family members—children, adolescents and adults—by no means believe they have been brainwashed; rather, they believe they have arrived at their own conclusions of their own volition; they believe they are appropriately responsive to evidence, as presented by their elders, by scripture, and by their own judgment. How, then, can Fischer and Ravizza distinguish cases of hypnosis from cases of ordinary enculteration? In response to this concern, Fischer and Ravizza assert that "it is tolerably clear that ordinary practical reasoning is in some way interestingly different in kind from practical reasoning in which crucial inputs have been implanted through direct manipulation, or the inputs are being processed through direct electronic manipulation."[141] Tolerably clear, perhaps, to them. But without specifying what precisely distinguishes the "induced inputs" of an external mechanism from the enculturated "non-induced inputs" that factor into one's "natural" course of practical reasoning, it seems they cannot sufficiently

[140] https://en.wikipedia.org/wiki/The_Most_Hated_Family_in_America.
[141] Fischer and Ravizza offer this response to Bratman, from "Replies," *Philosophy and Phenomenological Research*, Vol. 61, No. 2 (Sep., 2000), p. 476.

distinguish cases of a person genuinely "owning" his reasons-response mechanism from cases where such efforts are defective or bogus.[142]

I suggest, further, that Fischer and Ravizza's failure to distinguish such cases underscores at least two deeper problems with their compatibilist account. First of all, recall, as Fisher and Ravizza have conceded, that one's subjective "assumption of responsibility" does not *automatically* render one responsible. For, again, the very assumption of responsibility for one's reasons response mechanism may be informed by a reasons response mechanism that has been "implanted" in an inappropriate way. Yet, as suggested above, this appears to introduce a vicious circularity: one's efforts to assume ownership of one's mechanism will be undermined by the very mechanism one invokes to assume such ownership. Further, insofar as an individual cannot *produce* his own reasons response mechanism—as Fischer and Ravizza affirm, and as I argued in the previous section—it seems erroneous to hold an individual responsible for his reasons-response mechanism, *whether or not he correctly assumes ownership of it*. Indeed, as Smilansky has argued, "[o]ur decisions, even as ideal compatibilist agents, reflect the way we were formed, and we have had no opportunity to have been formed differently" (2003, p. 268). As such, Smilansky continues, Fischer

[142] As Zimmerman has put it: "[T]he distinction between evidence-sensitive and merely causally induced patterns of attitude-acquisition lies at the center of any genuinely source-historicist constraint on the development of autonomous agency from heteronymous beginnings." David Zimmerman, "Reason-Responsiveness and Ownership-of-Agency: Fischer and Ravizza's Historicist Theory of Responsibility", *Journal of Ethics*, Vol. 6, No. 3 (2002), p. 214. See a similar iteration of this argument developed by Todd R. Long, in "Moderate Reasons-Responsiveness, Moral Responsibility, and Manipulation" in *Freedom and Determinism*, (ed.) Joe Keim-Campbell, Michael O'Rourke, and Savie Shier (MIT Press, 2004).

and Ravizza's account "cannot form a sustainable barrier, either normatively or metaphysically, that will block the incompatibilist's further inquiries, about all of the central notions: opportunity, blameworthiness, desert, fairness and justice" (2003, p. 267).[143] So, yes, we can certainly *assum*e responsibility, as Fischer and Ravizza recommend, but, *the very assumption* does not supersede the conditions that have given rise to our reasons-response mechanism, given rise to our desire to assume responsibility for it, and given rise to very act of assumption, itself. Indeed, Fischer and Ravizza concede these difficulties and admit that they are "not offering a knockdown argument for the compatibility of causal determinism and taking responsibility" (1998, p. 236).[144]

4.4.1 Baker's Reflective Endorsement

In her essay, *Moral Responsibility Without Libertarianism* (2006), Lynne Rudder Baker provides a similar compatibilist account of agency and offers a suggestion that addresses Fischer and Ravizza's "ownership" dilemma. To develop her view, Baker invokes Frankfurt's above-described notion of hierarchical freedom and contends that "one important feature of Frankfurt's view of the hierarchical will that has gone largely unremarked is that it requires that the agent have a first-person perspective. A person must be able to conceive of her desires as her own—from the first-person—if she is to desire to have a certain

[143] Smilansky continues: "It is unfair to blame a person for something not ultimately under her control, and given the absence of libertarian free will, ultimately nothing can be under our control." (2003, p. 268)

[144] For further discussion of Fischer and Ravizza's "ownership" problem with respect to manipulation cases, see McKenna (2000, p.104). See also Mele (2000).

desire"(2006, p. 315).[145] Baker spells out her Reflective-Endorsement view as follows:

> (RE) A person S is morally responsible for a choice or action X if X occurs and:
> (i) S wills X
> (ii) S wants that she*[146] will X [i.e., S wants to will X].
> (iii) S wills X because she* wants to will X, and
> (iv) S would still have wanted to will X even if she had known the provenance[147] of her* wanting to will X

Baker summarizes her view as follows: "If I can say, 'These desires reflect who I am, and this is the kind of person that I want to be,' then (surely!) I am morally responsible for acting on those desires—whether determinism is true or not" (2006, p.318).[148]

How will RE allow Baker to address the "ownership" problem that arose in Fischer and Ravizza manipulation cases? Here, Baker appeals most directly[149] to the specifications of condition four, that is, S will still be responsible for her

---

[145] Baker claims, further, with Frankfurt, that "a first-person perspective is the defining characteristic of persons" (p. 315).

[146] Baker uses a * to denote "the agent" as the agent conceives herself from the first-person perspective. As noted, I have used quotation marks to establish the distinction.

[147] The term "provenance" is shorthand for a two-fold condition: (i) S knows that her* wanting to will X has causal antecedents that trace back to factors beyond her* control, and (ii) S knows of the causal antecedents that trace back to factors behind her* control that they are in the causal history of her* wanting to will X and that they are beyond her* control.

[148] See a similar articulation in Frankfurt (1988, p. 24): "Suppose that a person has done what he wanted to so, that he did it because he wanted to do it, and that the will by which he was moved when he did it was his will because it was the will he wanted. Then he did it freely and of his own free will."

[149] Baker also contends that a neuroscientist could not program into a person or a machine a first person perspective. As she puts it, a "first-person perspective cannot be acquired by neural manipulation "(p. 316; see also p. 322). See Baker (1981) for further arguments against the possibility that computers could acquire a first-person perspective. She has recently abandoned this idea (personal communication).

action X assuming she *would have* wanted to will X even if she had known the provenance of her* wanting to will X.[150] That is, according to (iv), Hypnotized Bill would *still* satisfy condition (iv) of RE just so long as (a) Bill knew he had been thus programmed and (b) having recognized his programming, Bill still would have wanted to will X. Indeed, to head off further objections, Baker invokes her "Completeness Clause," stipulating that an agent will still be responsible for her endorsement of P as long as "[t]here is no further knowledge of the circumstances of the agent's endorsement of his willing [P] that would lead the agent to repudiate his endorsement of his willing [P]." (2006, p. 318)

On the face of it, Baker's "Completeness Clause" does seem to address the ownership dilemma. After all, so long as an agent knew *everything* that significantly contributed to his or her decision making process, no epistemic deficiency *could* undermine the legitimacy of his or her assumption of responsibility. Yet her proposal raises some concerns. First, I suggest that her epistemic requirements are unreasonably high. As Nagel says, "The mind's work is never done" (1986, p. 129). That is to say: how might an agent possibly know at t1 whether there might exist further knowledge of the circumstances of her endorsement of p that, at t2, would lead her to repudiate her endorsement of p? Indeed, imagine someone doing everything she could to "cover all her bases," and affirming that no further revelation could possibly change her mind, only later to discover some crucial facts about herself or circumstances that would have materially factored into her considerations. Given that she had done her due

[150] Baker's response is in fact motivated by an argument developed by Derk Pereboom (2008: pp. 112-116).

diligence, should she then be let off the hook for her earlier endorsement? Indeed, recall the children of the Westboro Baptist church, who in fact confidently boast that "these desires reflect who I am, and this is the kind of person that I want to be, and no amount of further research could possibly permit me to change my mind!" Yet, in spite of such a child's passionate attestations, given the extent of her brainwashing, it seems unfair to view her as responsible for her attitudes and actions.[151] Indeed, imagine such a child growing up, and coming to realize she had been thus brainwashed. Should she blame her younger self for the views she held? Again, our intuitions seem to rule this out. As such, given our own necessarily limited perspective, RE seems to warrant at best a conditional attribution of blame.[152]

4.4.2 Moran's Avowals

Finally, let us revisit Moran's account of "avowals". Recall that Moran contrasts two stances a person can assume toward his own behavior or thought process: the spectator's *theoretical* or *descriptive* stance and the agent's *deliberative* or *first-personal* stance. And recall the case of Martha —a

[151] One might very well hold her "accountable," but not "ultimately responsible." I will discuss the distinction in chapter 7 and 8.

[152] See Nagel's discussion of this very problem in 1996, p. 126-134 and his practical solution. He writes: "[Human beings] want to be able to stand back from the motives and reasons and values that influence their choices, and submit to them only if they are acceptable. Since we can't act in the light of everything about ourselves, the best we can do is to try to live in a way that wouldn't have to be revised in light of anything more that could be known about us" (p. 127). And yet, again Nagel recognizes that "however much we expand our objective view of ourselves, something will remain beyond the possibility of explicit acceptance or rejection, because we cannot get entirely outside ourselves…" (p. 128).

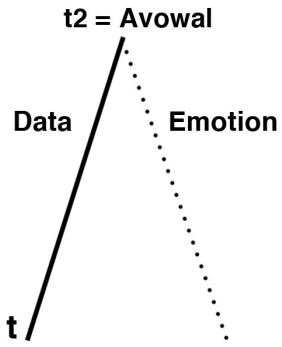
therapeutic patient who has been informed of the fact that she harbors deep hostility for her abusive father. Now, Moran suggests that Martha "may not doubt this. But without her capacity to endorse or withhold endorsement from [her belief], and without the exercise of that capacity making a difference to what she feels, this information may as well be about some other person, or about voices in her head" (2001, p.93). Thus, according to Moran, to "avow" an attitude requires *more* than that a person merely intellectually recognize that the attitude in question is one she possesses; rather, she must make an active, first-personal commitment to the attitude in question such that it makes a felt difference. And, according to Moran, it is up to a person whether she makes such a commitment or withholds it. Indeed, as he puts it, "in both the case of actions and attitudes, self-consciousness makes a difference to what the person's responsibilities and capacities are, with respect to his involvement in their development" (2000, p.32).

Now, in Section 3.33 I acknowledged that an avowal will effectively allow a person to more "fully own" a particular attitude or action. As such, if our patient, Martha, were to avow and more fully commit to her hatred of her father, I concede that such an avowal would allow for a fuller integration of her feelings, and that this would indeed be of therapeutic value. And here, too, I will acknowledge that such avowals do require an active commitment and that such a commitment will make a difference. Indeed, note that whenever we make a confession, or issue a genuine apology, or utter the words "I do," at a wedding ceremony, our utterances very often make a profound and felt difference. The question I wish to raise is whether our very commitment—that is, the actual

"merging" of our theoretical recognition of the truth of p, and our emotional "owning up to" the truth p is really "up to" our self-conscious agency, as Moran asserts.

To help clarify the issue, let us more fully flesh out the case of Martha. Let us suppose that Martha has sought psychiatric help because she feels "emotionally dead." During each session, Martha's therapist compiles a profile of Martha, filling it out with theoretical data, such as: 1. Martha was born in 1967 to a mother and an alcoholic father; 2. Martha's favorite color is green; 3. Martha was sexually abused by her father; 4. Martha harbors feelings of guilt and anger due to her father's behavior. During her therapy sessions, Martha registers all of these items of data from a position of intellectual neutrality. She merely "logs" them, as distinct bits of information. Now, given such detached neutrality, Martha's recognition of the facts of her case will be of negligible therapeutic value, for again, as Moran says, if Martha "become[s] aware of it only because [she] fully believes the interpretation given by [her] analyst, the attitude does not thereby become a conscious one. There is still work to be done" (2001, p.30). But the question is: *how* will Martha make the crucial transition from her theoretical, intellectual recognition of the data *to* a first-person felt commitment to the data in such a way that will make a "felt difference"?

To clarify the matter, consider following diagram:



At time t Martha regards the data: "I harbor guilt and anger for my father" as simply an item of neutral data, of no more importance than, "My favorite color is green." Martha cannot emotionally "connect" to the data because she is cut off from her emotions (hence the dotted line). Yet by t2, the theoretical data and Martha's dislocated emotions have linked up, such that Martha can now, so to speak, *own up* to the fact: "I was molested by my father" in a first-personal, felt way that will be of therapeutic value. But how exactly will this linking-up take place?

First, let us consider three possibilities, which, on Moran's account, we must *rule out*. We must rule out (1) the possibility that Martha's cognition of data (D) will reach a critical mass, and at that point *force* Martha to t2. We must also rule out (2) the possibility that an unexpected emotional upsurge (E) will *force* her to t2. And we must rule out (3) an automatic or causal convergence of the mass of data, on one side, and the emotional charge on the other (D+E), such that the combination of these two forces *compels* Martha to engage in a first-personal way with the data. We must rule out 1-3 because to attribute the transition from t to t2 to either a critical mass of theoretical data or to a spontaneous emotional upsurge,

or to an automatic or causal link up between the two, would be to suggest that such a transition would not be "up to" Martha in the sense Moran specifies, but, rather, that it would just "happen to" her.

So what needs to be the case such that Martha's transition can be understood as something that is "up to her" in the way Moran suggests? Again, Moran's answer to this question is of the familiar Rationalist variety: Moran suggests that the self-conscious rational agent can *mediate* between D and E. That is to say, on Moran's account, "Martha," as a self-conscious subject, occupies a position that is distinct from D and E, such that *she herself* (not D or E) can link up D and E. And yet, in light of the considerations offered above, it is difficult to understand how she can affect such a transition.

After all, given the conditions of Martha's case, we must rule out the possibility that she can affect the transition by means of purely rational analysis—for rational analysis alone will be insufficient to mobilize her emotional commitment. And, for the reasons discussed above, we should also rule out the possibility of her doing so by means of brute decision; that is, it seems implausible to suggest that Martha can "just decide" to link up D and E. Indeed, we may more clearly appreciate the implausibility of such a proposal by means of a simple consideration of the facts of Martha's case. That is, recall that Martha *already acknowledges* her belief that she was betrayed by her father (as mentioned, she accepts this as one of many theoretical facts about her), but, for any number of reasons—perhaps the workings of her unacknowledged fears and defense mechanisms—she has thus far been *unable* to own up to or avow her

belief in a way that makes a felt difference. As such, the very fact that Martha has been unable spontaneously to avow her belief in the desired fashion argues in favor of the proposition that the matter is not simply a question of brute decision. To thus suggest that Martha could, if she wanted, in a single bound, "leap over" the distance between her desire and her ability to avow her belief would be to deny the functional relevance of all the factors that have contributed to her existing condition. Moreover, Moran *himself* rejects the possibility of such voluntarism, claiming that "[t]he agency a person exercises with respect to his beliefs and other attitudes is obviously not like that of overt basic actions like reaching for a glass" (2001, p.114), and, further, that "the person's role in forming his attitudes is not to invoke a kind of willful or wishful capacity for self-creation" (2001, p.63). Yet, absent such a capacity, it is difficult to understand how Martha will self-consciously bring about required commitment. Indeed, I will suggest that for such reasons, Martha's commitment will not be something that is "up to" her self-conscious mediation, as Moran alleges. Rather, the connection will take place—bringing about her commitment—when, over the course of her therapy, her overall system becomes "ready" to do so, reaches a "critical mass" and the convergence automatically takes place.

4.5 Phenomenology and Time

I have thus far argued that, contrary to the efforts of the "I" to author, assume ownership of, or even commit to, the facts of the "me", in each case, the behavior of the "I" is governed by the facts of the "me". I recognize that these proposals

may contradict our intuitions. After all, few features of our human experience seem more inescapably obvious than that we *are*, in fact, single-party leaders of our deliberative activity, "determining" the course and outcome of our deliberations. As such, recall Kant's assertion that we *must* act "under the idea" of freedom, and Korsgaard's claim that "we *must* regard ourselves as agents, that being our situation, and not negotiable, for to be human is to have no choice but to choose" (2009, p.87, emphasis mine).[153] As such, this "sense" of freedom strikes us as an inalienable condition of our human predicament, and persuades us, perhaps more than anything, that I am behaving just as I think I am. That said, in this section, I am going to push for an even more counterintuitive and perhaps unpalatable claim. That is, I will argue that if we closely examine our phenomenology of deliberation—especially under time constraints—our phenomenology *actually does not* bear out our conception of ourselves as autonomous agents.

I thus offer *Briefcase*—an incident from my life. One sunny morning I left my San Francisco apartment at 8:30 to catch an 8:40 bus that would get me to work by 9:00. Arriving at the bus stop, and taking my seat on one of the orange, foldout benches, I noticed an unattended, expensive-looking briefcase perched against the front door of my apartment building. The moment I caught sight the briefcase, a

[153] See also Nagel (1997, p. 118): "We cannot evade our freedom. Once we have developed the capacity to recognize our own desires and motives, we are faced with the choice of whether to act as they incline us to act, and in facing that choice we are inevitably faced with an evaluative question. Even if we refuse to think about it, that refusal can itself be evaluated. In this sense I believe Kant was right: The applicability to us of moral concepts is the consequence of our freedom—freedom that comes from the ability to see ourselves objectively, through the new choices which that ability forces on us" See also, Sartre (2004, p. 350 and 360).

flood of thoughts began to course through my mind. The thoughts might be represented as follows: 1. How has it happened that this fancy briefcase ended up in this position? 2. Has perhaps a forgetful person—maybe even one of my neighbors—left it there? 3. Might this person be rushing back to the apartment building right now to retrieve the briefcase? 4. Could it be that someone found or even stole the briefcase and set it down there, hoping the owner would retrieve it? 5. Has a friend left the briefcase there for someone else to pick up (the building had no buzzer)? 6. But who would possibly agree to such a crazy arrangement? 7. If any of possibilities 2-5 are correct, and someone is in need of the briefcase, should I rush up and grab it? 8. After all, if I don't retrieve the briefcase, and do it right away, someone else, and probably someone less conscientious, is bound to grab it. 9. Now I can see the bus coming up the street. I must make up my mind very quickly about whether to get the briefcase. 10. If I miss my bus I could take a taxi to work, but it is rush hour and I might not be able to flag a taxi in time. Is it worth the risk? 11. Why haven't I grabbed the briefcase yet? Why have I let so much time elapse? And what does this say about me? Am I inconsiderate? Am I lazy? Am I irresponsible? Why haven't I felt sufficiently motivated to retrieve the briefcase? 12. Why am I wasting crucial time worrying about what my waste of time says about me? 13. Could there be a bomb in the briefcase? 14. Could this be some strange trick or prank? 15. After all, really, what are the chances of someone either accidentally leaving a briefcase out on the sidewalk like this, or leaving it on the sidewalk for the reasons mentioned? 16. But, then again, what's the probability of this being a prank? 17. Am I just trying to talk myself out of

retrieving the briefcase because I don't want to deal with it? 18. If I don't grab the briefcase, will I regret not having done so? 19. Suppose it was my briefcase, wouldn't I hope that a responsible person such as myself would grab it? 20. But now the bus is one block away. If I run across the street to retrieve the briefcase I will almost certainly miss my bus and, in all probability, will be late for work. Is it worth the risk? 21. Aw, hell, I might as well run the risk; I'll go grab the briefcase. 22. I go and grab the briefcase.[154]

Let us consider just a few of the particular features of my thought process. To begin with, notice I did not *choose to engage* in the process itself. Rather, the process "took me over," and, so to speak, "threw me downstream," the moment I set eyes on the briefcase. Further, notice that questions concerning the briefcase engaged my attention due to my natural susceptibilities to circumstances like this; for example, my disposition to be helpful, especially when it comes to things like lost wallets, keys and briefcases, especially as I myself have often lost such items and strangers have returned them to me. Of course I had hoped to pass my time at the bus stop otherwise engaged—perhaps reading the novel I had brought. But I *could not control the extent to which the briefcase-related thoughts commanded my attention and coursed through my mind*. Notice, further, that as my thoughts raced, I was cognizant of the fact that, for every moment I did not run across the street to retrieve the briefcase, I left open the possibility that someone else—someone less conscientious—could do so, and I chastised myself for the fact that I had not run across the street and retrieved the briefcase. Yet even as I fretted over

[154] It turned out, incidentally, that someone had lost the briefcase; someone had found it and then just set it up against the doorstep of my apartment building. The owner, who did not live in my building, was very grateful that I returned it.

the fact that I had not taken action, I was unable to regulate the force, or the "valences," of the conflicting voices in my head. Indeed, it became quite clear to me that while there were "warring voices" in my head, and while all of these voices genuinely expressed my concerns, there was no "I"—that is, no "single impartial party" no unitary, transcendent "I", *overseeing* and *conducting* this process. As Dennett has put it, I could not act as "parade marshal for the queue of considerations-to-be reviewed, putting each in its proper place in line."[155] Indeed, this "parade" advanced according to "its own" predilections, at its own speed, and in whatever direction it chose. As such, while there was no question that the thoughts coursing through my head were "mine," at the same time it was anguishingly clear that "I"—a dislocated agent—was not "in control" or, indeed, even "actively thinking" them.

Now, one may contend that instances such as this are exceptional, that, generally speaking, when we deliberate, we are not subject to such temporal pressures, and that, as such, under ordinary circumstances, we correctly experience ourselves as "taking control" of the process. Yet we may account for such contrasting phenomenology in a number of ways.[156] First, note that when we are very strongly inclined in one direction or another, we do not experience a tension or a schism among our competing beliefs or desires and as a result we do not experience a sensation of internal conflict or fragmentation. We do not take ourselves to be at the mercy of, or pushed around by, competing thoughts or desires. Rather, in virtue of the "univocality" or alignment of our disposition we

[155] Daniel Dennett, *Elbow Room, The Varieties of Will Worth Wanting* (Cambridge: Cambridge University Press), 1985, p. 86.
[156] I will further develop these ideas in Chapter 7.

may, indeed, feel like a "single party" determining our course. Indeed, under such circumstances, as Frankfurt has put it, our conviction may "resound endlessly" through us (1998, p. 21). Second, when we have the luxury of more time, and can more patiently consider our options, we naturally feel more at ease, less rushed and more in control. Yet this does not suggest that we are exercising supervisory control over the order or the intensity of our competing considerations; they are just progressing at a more leisurely rate and are attended by an overall feeling of comfort or ease.

4.6 Agency Conclusion

Note that the central focus of this chapter has *not* been the question of whether or not we engage a kind of "deep" agency unavailable to creatures lacking in self-consciousness. This has never been in dispute. Rather, my focus has been the Rationalist characterization of two central features of deep agency—the Second-Order Component and the Responsibility Component. I have not challenged the second-order component. I take it as incontestable that persons reflect upon, and attend to, themselves—their attitudes and actions. Rather, I have challenged the second, "Responsibility" component of deep agency, that is, the Rationalist proposal that the "I" can attend to the facts of the "me" in a way that does not derive, or flow from, the facts of the "me"—in a way, that is, where the "I" can "unilaterally" influence the facts of the "me." I have argued to the contrary, suggesting that the behavior of the "I" is constrained by a vast network of personal facts—constrained, that is, by antecedent desires, dispositions,

preferences, truth-tracking mechanisms, etc. Indeed, I have even militated against what I take to be the lynchpin of the Rationalist position, so forcefully emphasized by Descartes, that it is a "truth which nobody should deny, that there is nothing entirely in our power except our thoughts". Indeed, I have argued that while it *feels* as if we are generating our thoughts, commanding them, or "employing" them such as one may put to use a fork and a knife, such actually is not the case. Again, as Arpaly has put it: "the mere first person *experience* of [one] having control over [one's] mental life is not by itself a surefire indication that [one] *actually* has control over [one's] mental life, in any meaningful sense of 'self-control'" (Arpaly, 2003: 19). Indeed, as Nietzsche, following Schopenhauer, has said, "a thought comes when 'it' wants, and not when 'I' want. It is, therefore a *falsification* of the facts to say that the subject 'I' is the condition of the predicate 'think'" (2002, p. 17, emphasis his).[157]

Yet perhaps we have overrated the importance of the kind of responsibility ostensibly licensed by deep agency. Perhaps, at least when it comes to making a difference with respect to ourselves, improving ourselves, regulating our behavior, and pursing our projects, we can dispense with the autonomous "I". Perhaps, when it comes to such activities, an "I" that is empirically situated is all we really need. Such will be the subject of the following chapter.

[157] See Schopenhauer: "Thoughts come not when *we* want them, but when *they* want to." *Parerga and Paralipomena* Vol. 2 (trans. E.F. J. Payne) (Clarendon Press: Oxford, 1974) par. 37 p. 51.

CHAPTER 5

SELF-COMPOSITION

5.1 Introduction

In Chapter 3 I explored the question of whether the "I," in virtue of its distance from the facts of the "me", could bring a unique repertoire of epistemic powers to bear on these facts. In Chapter 4 I explored the question of whether the "I," in virtue of this same distance, could produce or assume ownership of the facts of the "me" in a way that was not entirely parasitic on the facts of the "me". In both cases, I argued that whatever powers the "I" brought to bear on the facts of the "me" overwhelmingly derived from the facts of the "me". That said, I did *not* rule out the possibility that I *could* achieve self-knowledge; *nor* did I reject the proposition that I *can* take action upon my facts, however such "taking action" is to be understood.

For some, such powers, unique to self-conscious agents, might be regarded as fully adequate, *all one really needs* to pursue one's projects. After all, we very commonly engage in self-reflection, come to realize that we do not approve of some personal quality, perhaps an aspect of our physical appearance, our character, or our skill-set; and we certainly can take concrete steps to improve in these departments. We can go on diets, visit therapists, hire coaches, meditate, develop means of controlling our temper, etc. We often fail to achieve our objectives, but sometimes we succeed, and sometimes we succeed to a

phenomenal degree, and when we do so, our achievement can give rise to feelings of thrilling self-empowerment, assuring us that "anything is possible," that "if there's a will there's a way," that, indeed, we are self-creating Masters of ourselves and of our destinies.

Such optimism perhaps finds its most rousing endorsement in Sartre's existential dictum, *l'existence précède l'essence*, the proposition that a human being does not arrive with a "nature" that determines or otherwise constrains his personal development, and that, to such an extent, "there is no explaining things away by reference to a fixed and given human nature".[158] Rather, on Sartre's view, it is *entirely in our power*, and, indeed, it is our *responsibility*, to shape ourselves into the kind of people we wish to become. Korsgaard has endorsed a kindred view, affirming that "[t]he form of the human is precisely the form of the animal that must create its own form" (2009, p. 130). And, she has asserted with characteristic brio: "As a rational being, as a rational agent, you are faced with the task of making something of yourself, and you must regard yourself as a success or a failure insofar as you succeed or fail at this task" (2009, p. xii).[159]

Again: "[Y]ou must regard yourself as a success or a failure insofar as you succeed or fail at this task." This, indeed, is a formidable prospect, enormously encouraging or oppressive, depending on one's bent. Yet one may wonder how

[158] The whole passage reads: "If existence really does precede essence, there is no explaining things away by reference to a fixed and given human nature" (157, p. 356).

[159] Nietzsche offers a similar exhortation, recommending that we must "survey all the strengths and weaknesses of [our] nature and then fit them into an artistic plan until every one of them appears as art and reason and even weaknesses delight the eye." Nietzsche, (2001) §290, p. 163.

much self-determining influence one can reasonably be expected to shoulder. Unsurprisingly, I will assume a critical stance here, arguing, as I have before, that such powers are in fact rigidly constrained and, as such, that one should only *very provisionally* hold oneself responsible for the success or failure of one's self-projects. I will proceed as follows. First, I will examine our means of *determining* or *emending* our existing attributes, what I shall call acts of "Self-Composition". Here, I will briefly revisit ideas of early Frankfurt and Schechtman and then move on to Frankfurt's notion of "volitional necessities". Following this, I will explore the means by which, according to Korsgaard, we establish our "integrity," organizing ourselves, or "pulling the parts of ourselves together," so as to become fully integrated, or "united" persons.[160] I will refer to such endeavors as acts of "Self-Integration". In all such cases, I will argue that, while we can indeed exercise a good deal of influence over our facts, the extent to which we can do so, our very means of doing so, and even our desire to do so, derive from and are constrained by our facts, and, further, that what we ultimately become has little to do with our self-conscious, deliberate efforts. As such, I will endorse Arpaly's claim that "[i]t is the exception, rather then the rule, that a person's character is substantially self-made, which is why a self-made good character is so impressive in the first place." (2003: 141-142)

5.2 Self-Composition - Early Frankfurt and Schechtman

Recall Frankfurt's "hierarchical" conception of self-identification. Here, one evaluates a first-order desire and develops a second-order desire to either endorse

[160] This, as well, will involve some review.

or reject this desire. If one decides to endorse the first-order desire, one thereby "identifies" with it and, in doing so, "makes [it] more truly his own". Or else one can "withdraw from," and therefore "externalize," or "alienate" a desire, in such a way as to render this desire *no longer* his own (or, at least, *less truly* his own). As Frankfurt tells us "[i]t is these acts of ordering and rejection—integration and separation—that create a self out of the raw materials of inner life." Indeed, Frankfurt continues, these acts "define the intrapsychic constraints and boundaries with respect to which a person's autonomy may be threatened even by his own desire" (1988, p. 170). That is, according to Frankfurt, when we undertake this process of "ordering and rejection" we are literally *creating* our identities, determining which desires or beliefs truly represent us.

Yet, as discussed in section 3.31, Frankfurt's account of identification-as-self-construction is vulnerable to a number of objections. In the first place, recall that identification with a particular desire will by no means *guarantee* the self-representational authority of that desire. I may, for example, wish to desire someone romantically, yet the mere desire to desire her as such by no means renders it the case that I truly do, or that I even *could*, desire her, as such. Nor will the externalization of a desire necessarily verify its non-representational status. That is, my desire to externalize or alienate my hostility for my father by no means guarantees that such hostility is any less self-representative, any less my "own".[161] As such, in both cases, regardless of my hopes and intentions, my acts of "identification" may not in any way establish "intrapsychic constraints and

[161] See Velleman's treatment of Freud's "Rat-Man" case in "Identification and Identity," from Buss (2002).

boundaries". Indeed, as discussed earlier, even when one regards one's commitment as "wholehearted," the self-representational authority of one's desire may be misleading, an expression of fantasy or wishful thinking.[162]

We encountered similar ideas in Schechtman's narrative account. Here, recall that, according to Schechtman, a person "creates his identity" through an act of interpretive self-construction, "forming an autobiographical narrative—a story of his life." One figuratively or literally composes a "conventional, linear narrative" (1996, p. 96) where "the incidents that make up his life are not viewed in isolation, but interpreted as part of the ongoing story that gives them their significance" (1996, p. 97). As time passes, persons weave their more dramatic, or salient, events into a meaningful narrative, or autobiography, according to which they can explain, for example, why they behaved, as they did, why they felt as they did, and how they became the people they are today. Indeed, Schechtman regards this activity as a *necessary* component of creating and sustaining our personhood. [163] Yet, according to Schechtman, not *any* story will do. Rather, in order for one to *responsibly* engage in narrative self-construction, one must respect what she calls "reality constraints"[164]. As she puts it, "[a] narrative that reveals the narrator to be deeply out of touch with reality is thus undermining of

[162] As Velleman puts it: "When Frankfurt describes us as identifying with some of our motives and alienating others, his description rings true, I suspect, because it accurately describes this common defensive fantasy [of identifying with just one aspect of our being such as our love for our parents]. We do indeed identify with some of our motives, but we thereby engage not in self-definition but self-deception. We identify with some of our motives by imagining ourselves as being those motives, to the exclusion of whatever might complicate or conflict with them." (2002, p. 109)

[163] See Strawson's (2004) objections to Schechtman's narrative account.

[164] As she puts it, such constraints must "fundamentally cohere with reality" (1996, p. 119).

personhood and hence cannot—at least with respect to those elements of the narrative which seem grossly inaccurate—be identity-constituting" (1996, p. 120). Such reality constraints require one to recognize "matters of observable fact," and matters of "factual interpretation" (1996, p. 125-126). One, for example, will need to properly recognize facts of past experiences (one was in fact born in a specific country and on a particular date; one did, in fact, attend a particular school, and develop measles); and one's interpretation of factual observation must obey certain standards of plausible inference (paranoiac delusions and conspiracy theories, for example, are ruled out). Further, one's self-conception must respect one's "robust inclinations"—enduring propensities or preferences (a short-temper; independent thinking; romantic fidelity; punctuality) that must be "relatively stable, coherent and powerful" (2004, p. 415).[165] Indeed, commenting on Frankfurt's views on identification, Schechtman contends that one's robust inclinations "are to be given presumptive authority even when we do not identify with them, and so the threshold for excluding them will be higher" (Schechtman, 2004, 426). That is, Schechtman's approach places less emphasis on "settling" the tensions introduced by our conflicting inclinations, in favor of "establishing safe boundaries within which these conflicts can be allowed to play themselves out" (2004, p. 426). Thus, on Schechtman's account, to compose a responsible narrative requires that one hew both to objective facts about the external world and to hardwired facts of one's own nature.

[165] Schechtman (2004, p. 415). Schechtman provides, in addition, that such inclinations must "not have their origin in an obvious physical or psychological pathology." (2004, p. 415).

5.2.1 Frankfurt's Volitional Necessities

That one's self-constructive efforts must obey a set of core characteristics or "robust inclinations" is underscored in Frankfurt's discussion of "volitional necessities". To set the stage, Frankfurt asserts that:

> [t]he idea that the identity of a thing is to be understood in terms of conditions that are essential for its existence is one of the oldest and most compelling of the philosophical principles that guide our efforts to clarify our thought. To grasp what a thing is, we must grasp its essence—viz., those characteristics without which it is not possible for it to be what it is. Thus, the notions of necessity and identity are intimately related. (Frankfurt 1999, p. 113)

Indeed, Frankfurt claims, *pace* Sartre, that a person's character is *no exception* to this rule, that it in part obeys set of inviolable "volitional necessities"—cares or desires that one not only *cannot desire* to violate, but that, for him, are "unthinkable" (1988, p. 187). As Frankfurt puts it: "Our essential natures as individuals are constituted, accordingly, by what we cannot help caring about. The necessities of love, and their relative order or intensity, define our volitional boundaries. They mark our volitional limits, and thus they delineate our shapes as persons" (1999, p. 138). Thus we may imagine, as Frankfurt does, a woman who has agreed to put her child up for adoption, and who may believe she wishes to do so, but who "cannot bring herself" to go through with it. Or we may imagine, again as Frankfurt does, a soldier who, in spite of his most thoroughgoing training, is unable to bring himself to execute orders to discharge nuclear weapons. Frankfurt explains that "[h]e cannot perform any of them, because he is prevented by a volitional constraint; that is, he cannot will to perform them. Even though he may think it would be best for him to perform one of the actions, he

cannot bring himself to perform it. He cannot volitionally organize himself in the necessary way. If he attempts to do so, he runs up against the *limits of his will*" (1999, p. 111, emphasis his).[166] In such cases, again, involuntary cares are resistant to any oppositional forces, and, as such, qualify as definitive, inviolable, features of one's identity.[167]

5.3 Korsgaard on Self-Integration

With Frankfurt and Schechtman, Korsgaard also acknowledges that a person's effort at self-composition must respect fundamental constraints, or "internal standards," that is, standards that "[arise] from the nature of the object to which it applies, from the functional or teleological norms which make it the object that it is."[168] Yet Korsgaard's standards do not involve "volitional necessities" rooted in Frankfurtian cares, or "robust inclinations" of the Schechtman sort. Rather, Korsgaard suggests a human being can only render herself a "person," and can only determine what *kind* of "person" she is—by means of her norm-governing *practical reason*. As she puts it: "Human beings therefore have a distinct form of identity, a norm governed or practical form of identity, for which we are ourselves responsible" (2009, xi). Indeed, as we have seen, on Korsgaard's view, in order

[166] See these examples at (1999, p.111).

[167] As Frankfurt puts it elsewhere: "Now the character of a person's will constitutes what he most centrally is. Accordingly, the volitional necessities that bind a person identify what he cannot help being. They are in this respect analogues of the logical or conceptual necessities that define the essential nature of a triangle. Just as the essence of a triangle consists in what it must be, soothe essential nature of a person consists in what he must will. The boundaries of his will define his shape as a person." (1999, p.138)

[168] Korsgaard (2009, p. 14). See also, Korsgaard (1997, 215-254, especially pp. 249-250).

for a "person" to be able to perform actions that can be attributable to *her* as opposed to *her parts*, she must play an overarching, "monarchic" or "aristocratic" role with respect to her parts, "[pulling] the parts of the soul together into a unified system"[169] and thereby safeguarding her "volitional unity" Korsgaard (1999, p. 28)[170]. And she does this via the exercise of rational authority. That is, without exercising such rational, authoritative, self-command, Korsgaard warns that a person's behavior and, indeed, her very identity, will succumb to the force of her mutinous desires.

To make her case, Korsgaard offers the example of Jeremy, a "democratic" soul who lacks the authority of such an overarching, unifying, rational power, and whose life is thus "completely dependent on the accidental coherence of his desires":

> Jeremy, a college student, settles down at his desk one evening to study for an examination. Finding himself a little too restless to concentrate, he decides to take a walk in the fresh air. His walk takes him past a nearby bookstore, where the sight of an enticing title draws him in to look at the book. Before he finds it, however, he meets his friend Neil, who invites him to join some of the other kids at the bar next door for a beer. Jeremy decides he can afford to have just one, and goes with Neil to the bar. While waiting for his beer, however, he finds that the noise gives him a headache, and he decides to return home without every having the beer. He is now, however, in too much pain to study. So Jeremy doesn't study for his examination, hardly gets a walk, doesn't buy a book, and doesn't drink a beer. (1999, p. 19)

[169] Korsgaard (1999, p. 22)

[170] Korsgaard elaborates: "To be a thing, one thing, a unity, an entity; to be anything at all: in the metaphysical sense, that is what it means to have integrity. But we use the term for someone who lives up to his own standards. And that is because we think that living up to them is what makes him one, and so *what makes him a person at all*" (1996b, p. 101, emphasis added).

Korsgaard notes, further, that "it's only an accident that each of Jeremy's impulses leads him to an action which completely undercuts the satisfaction of the last one. But that's just the trouble, for it's only an accident if this does *not* happen. The democratic person has no resources for shaping his desires to prevent this, and so he is at the mercy of accident" (1999, p. 20). That is, a defective "democratic" person lacks an overarching rational captain to pull his parts together into a cohesive whole. On the other hand, a "monarchic" or "aristocratic" person can self-consciously "shape his desires" and thereby *render it the case* that his behaviors obey his over-arching, supervisory will. Indeed, were one to lack this ability, Korsgaard warns, one's desires could tear a person apart and open up a psychological "fault line" (1999, p. 20).[171]

Of course, for many of us, the prospect of exercising such an aristocratic power of "self-possession" whereby one may be "entirely self-governed, so that all of [one's] actions, in every circumstance of [one's] life, are really and fully [one's] own: never merely the manifestations of forces at work in [one] or on [one]…"[172] will be greatly appealing. Yet, I suggest that Korsgaard's aristocratic vision is vulnerable to several objections. To begin with, let us return to the case of Jeremy, this time assuming he does in fact possess an aristocratic, or monarchic, soul. How will his day proceed?

> Jeremy wakes up and writes a list of things he would like to do today. He'd like to study for three hours without interruption, take an hour's walk that ends up at a book store, buy a copy of Camus'

[171] Korsgaard elaborates that the defect of non-aristocratic characters "is like a geological fault line, a potential for disintegration that does not necessarily show up, and so long as it doesn't, these people have constitutional procedures and so they can act" (1999, p. 20).
[172] Korsgaard (1999, p. 20).

> *The Stranger* and finally have a beer with some friends at a bar. So Jeremy sits down and begins studying. He feels the inviting air outside and considers going for a walk now, but he suppresses the urge and finishes up his studying. Then he goes for a walk and considers stopping at a nearby bookstore instead of the one that is an hour away, but he overcomes the urge and continues on to the proper bookstore. As he is looking for his book, a friend asks him to join him right away at the bar. Jeremy politely declines, finds his book and only then joins his friend at his bar.

Jeremy thus appears to be in absolute command of his faculties; he is a man who cannot be tempted away from his resolutions by competing desires, who does everything that he, the ruler of his soul, sets out to do. He is, in short, the type of person many of us try, but almost invariably fail, to be. Why do we fail? For a helpful clue, consider the following case proposed by Paul Edwards[173].

> Let us suppose both A and B are compulsive and suffer intensely from their neuroses. Let us assume that there is a therapy that could help them, which could materially change their character structure, but that it takes a great deal of energy and courage to undertake the treatment. Let us assume that A has the necessary energy and courage while B lacks it. A undergoes the therapy and changes in the desired way. B just gets more and more compulsive and more and more miserable. Now it is true that A helped form his own later character. But his starting point, his desire to change, his energy and courage, were already there.

As Edwards points out, A's and B's "starting points," that is, their initial "energy and courage," crucially factored into the success of their respective undertakings. Now, I propose that one's ambition to wield aristocratic authority over one's desires is no exception. That is, some people possess motivation and strength of character or willpower sufficient to exercise such authority. Others—most—do

[173] Dennett (1985, p. 84) takes this passage from Edwards' essay "Hard and Soft Determinism" that originally appears in Hook (1961).

not.[174] Of course, a person who does not at first possess sufficient motivation and strength of character can surely attempt to cultivate sufficient motivation and strength of character. But notice that the very commitment to cultivate sufficient motivation and strength of character involves the same difficulty: that is, one will require sufficient motivation and strength of character to succeed in the endeavor to cultivate sufficient motivation and strength of character.[175] Of course if one could *voluntarily generate* sufficient motivation and strength of character, one could bring this regress to a halt. And yet, if one possessed such a capacity, one presumably would not have encountered the problem in the first place. Further, as discussed in Section 4.33, it seems highly unlikely that, through some sort of mental volition, I could generate enthusiasm for something for which I actually don't care or that I could will myself to love a person that I don't like—no matter how "reasonable" I deemed such desires. So, while we may be able to exercise limited control with respect to defying or suppressing our existing desires, I suggest that the efficacy of this controlling part will itself be constrained by underlying traits—elements over which we, in turn, can exercise only limited control. Thus, I suggest that our "aristocratic" self that attempts to "rule" over our parts will *just constitute another part of us*, and the role it plays in "pulling our parts together" will be no less contingent than the roles played by the other parts.

---

[174] See also Nietzsche (1997, p. 64): "[*T]hat* one *wants* to combat the vehemence of a drive at all, however, does not stand within our power nor does the choice of any particular method; nor does the success or failure of this method. What is clearly the case is that in this entire procedure our intellect is only the blind instrument of *another drive*…"

[175] On will as a finite resource, see *The Willpower Instinct*, Kelly McGonigal Ph.D. (New York: Penguin, 2012), especially p. 54-80.

To put it a bit differently, I contend that *we cannot exercise aristocratic control over the part of us that seeks to exercise aristocratic control*.[176]

What about Korsgaard's "fault-line" and the danger of self-disintegration? Korsgaard's suggestion seems to be that only an over-arching monarchic captain "pulling the parts of the soul together" can ensure that a person's character won't simply founder and deteriorate under the pressure of conflicting desires. Yet some have argued that Korsgaard exaggerates the threat of such self-decomposition. Schechtman, for example, attests that such accounts of self-integration are guilty of "over-estimating the danger of internal civil war" (Schechtman, 2004, p. 425-426). On Schechtman's view, our hardwired robust inclinations are perfectly capable of doing the work of holding ourselves together. Further, as Blackburn has put it, a well-integrated or harmonious person, like a harmonious ship, will realize its harmony *not* through the command of an overseeing unitary caption, but through the "fortunate composition of [her] crew members: these will be crews in which the propensities are to the gay, benevolent, temperate, industrious, cheerful, hopeful, resolute" (Blackburn1998, p. 245). Finally, Daniel Dennett, perhaps the most vocal critic of the overriding, unitary, Captain, puts it as follows:

> In our brains, there is a cobbled together collection of specialist brain circuits, which, thanks to a family of habits inculcated partly by culture and partly by individual self-exploration, conspire together to produce a more or less orderly, more or less effective, more or less well-designed virtual machine, the *Joycean* machine. By yoking these independently evolved specialist organs together in common cause, and thereby giving their union vastly enhanced powers, this virtual machine, this software of the brain, performs a sort of internal political miracle. It creates a virtual captain of the crew, without elevating any one of them to long-term dictatorial

[176] See another interesting analysis of this problem of "self-mastery" in Leiter (2007, esp. p. 14).

> power. Who's in charge? First one coalition and then another, shifting in ways that are not chaotic thanks to good meta-habits that tend to entrain coherent, purposeful sequences rather than an interminable helter-skelter power grab." (Dennett 1991, p. 228)

On Dennett's account, our sense of captain-hood is really an illusion, a "virtual" captain of crew, produced by a Humean[177] multiplicity of cognitive processes that "conspire together" to constitute our "more or less orderly, more or less effective, internal political machine." Such processes, just like our internal biological functions, are regulated not by "us"—self-conscious captains, residing over and above our biological and psychological faculties—but rather by a network of adaptive mechanisms that operate primarily beneath our conscious awareness. Note, importantly, that Dennett *does not deny* our powers of self-control and deliberation; he just suggests that the "self" driving and regulating these operations is not a unitary Executive Controller, but, rather, a multiplicity of parts working together "in ways that are not chaotic thanks to good meta-habits that tend to entrain coherent, purposeful sequences".[178]

Finally and briefly, one might argue that Korsgaard has unfairly coronated reason as the only means of uniting our disparate parts. True, the net of reason

---

[177] Hume: "I may venture to affirm of the rest of mankind, that they are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in a perpetual flux and movement" *(T*: 1.4.6, p. 165).

[178] As Dennett puts it elsewhere: "So wonderful is the organization of a termite colony that it seemed to some observers that each termite colony had to have a soul (Marais, 1937). We now understand that its organization is simply the result of a million semi-independent little agents, each itself an automaton, doing its thing. So wonderful is the organization of a human self that to many observers it has seemed that each human being had a soul, too: a benevolent Dictator ruling from Headquarters." Dennett (1991, p. 416) See also Damasio, chapter 9, "The Autobiographical Self" from *Self Comes to Mind* (New York: Random House, 2011)

casts widely over a range of our facts. But as the later Frankfurt has said, our passions also serve to unify the elements of our character, to constrain the will, and can often have "final say" with respect to determining our actions and making us the kind of people we are and wish to be. Indeed, as he puts it, it is our involuntary *caring* that can be "indispensably foundational as an activity that connects and binds us to ourselves. It is through caring that we provide ourselves with volitional continuity, and in that way constitute and participate in our own agency" (2004, p. 17).

5.4 Self-Composition – Conclusion

In Chapter 3, I did not reject the proposal that the self-conscious "I" could access a perspective on the facts of the "me". What I did argue was that the extent to which the "I" could gain such access and the accuracy of such knowledge was overwhelmingly dependent on the facts of the "me". In Chapter 4, I did not reject the proposal that the self-conscious "I" could act on or influence one's facts. What I did argue was that the behavior of the "I"—the extent and nature of its contribution and influence—was, again, wholly parasitic on the facts of the "me". Here, again, I am not disputing the proposition that we can go to work on ourselves, amend ourselves or hold ourselves together. Again, we can, to some degree, do all such things. Yet I have pushed for the same basic thesis; that is, the extent to which one can do so is entirely constrained by the facts of one's constitution. As such, in all three cases, I have suggested the behavior of the "I" is overwhelmingly constrained by the facts of the "me". That is, I have challenged

the proposition that the "I" can exert, as I put it, "unilateral influence" on the facts of the "me".

Yet recall that I identified the capacity of the "I" to exert "unilateral influence" on the facts of the "me" as an indispensable component, indeed, as the *very cornerstone*, of the Rationalist position. And recall, per the Rationalist account, that the "I" obtains such a capacity *in virtue of the reflective distance* it can ostensibly assume with respect to such facts. That is, it is in virtue of such distance that the "I," as Korsgaard puts it, is *no longer dominated* by such facts, and, as such, can *attain leverage* over them, intermediate among them, influence them, produce them.[179] Yet, I have contested that over a broad range of cases such unilateral or autonomous powers are nowhere to be found, that, in all such instances, the behavior of "I" can only be understood as an *expression* of such facts—indeed, that the "I" acts, as Arpaly puts it, as their "dupe". And yet, if such is the case, that is, if the autonomy of the "I" with respect to the facts of the "me" is illusory, we might likewise wonder whether any leverage-conferring *distance* actually stands between the "I" and the "me,"—whether, indeed, the "I" and the "me" should be recognized as distinct existences. Such will be the subject of the following chapter.

[179] Recall again Korsgaard's remark: "Our capacity to turn our attention on to our own mental activities is also a capacity to *distance ourselves from them*. I perceive, and I find myself with a powerful impulse to believe. But I back up and bring that impulse into view and then I have a certain distance. *Now the impulse doesn't dominate me* and now I have a problem."

CHAPTER 6

THE HOLISTIC MODEL – A CREW OF CAPTAINS

6.1 Introduction

The preceding three chapters have constituted the "negative" portion of this thesis, a critical assessment of central aspects of the Rationalist model of the self. I have (1) Presented a number of scenarios according to which, on Rationalist grounds, the "I" can bring a set of epistemic and agential powers to bear on the facts of the "me"—powers, that is, that do not derive from and are not constrained by the facts of the "me" and (2) I have disputed the characterizations of these scenarios in an effort to demonstrate that, in each case, the causal powers of the "I" do in fact derive from and are constrained by the facts of the "me". As such, I have objected to the proposition that the "I" can exercise "unilateral" causal influence over the facts of the "me". Yet, if such is the case, then one might wonder how to properly understand the *existential* relationship the "I" bears with respect to the "me". That is to say, if the behavior of the "I" must be understood as an expression of the facts of the "me"—if, *functionally*, the "I" and the "me" are inseparable—then to what extent can we justifiably regard the "I" and the "me" as *existentially* distinct? Or, to express the question a bit differently: Allowing that it is possible—pushing Arpaly's claim to the extreme—that the "I" *always* acts as a dupe of the "me," might we not justifiably infer that the "I" and the "me" are not distinct, at all? Such will be the central contention of this chapter, and, as such, will constitute what I regard as the "finishing blow" to the

Rationalist conception of the self. I will proceed as follows. First I will offer a brief historical overview of the sequestering and coronation of the captainly, rational "I," focusing on Plato, Descartes and Kant, and then culminate in a brief recap of our contemporary accounts. In reviewing this history I will focus on an enduring argument that invokes what I shall call the *Distinct Object Thesis*. Following this historical overview, I will present three models of the self intended to capture candidate metaphysical relationships shared by the "I" and the "me"—Unilateral, Complimentary and Holistic—and then explain my preference for the latter, Holistic, model. In the concluding two chapters, in an effort to more fully spell out and render palatable the Holistic model, I will develop its various metaphysical and ethical implications, arguing, ultimately, that the Holistic model offers a more accurate and ethically beneficial view of the self.

6.2 The Autonomous, Captainly "I"—A (Brief) Historical Overview[180]

(1) Plato

It is perhaps in Plato where we encounter the first detailed Western formulation of the Rationalist Story, so here will be a natural place to begin. In *Phaedrus*,[181] Plato compares the human soul to a chariot with two winged steeds. One "honorable," "upright and clean-limbed," white steed is the "spirited" part of the soul. This part is "a lover of glory, but with temperance and modesty," and

[180] This overview is obviously not meant to be comprehensive; I am just tracking some prominent historical strains of argument.

[181] See Plato's *Phaedrus*, (trans.) R. Hackforth (Cambridge: Cambridge University Press, 1972). The passages in the *Phaedrus* concerning the chariot analogy occur at 246a-247d and 253d-257b. My brief discussion of this allegory derives from Hackforth's commentary.

"needs no whip, being driven by the word of command alone" (253 d-e). The other, black, horse embodies the "appetitive" part of the soul; it is "crooked of frame, a massive jumble of a creature, hot-blooded, consorting with wontonness and vainglory, shaggy of ear, deaf, and hard to control with whip and goad." Finally, the driver personifies Reason, the "reflective or calculative part of the soul"[182] that must master the dark and lustful steed that threatens to break loose and run amok.

Yet it is in the *Republic* where we first encounter a direct reference to a *ship captain*, here described as the philosopher or "star gazer" (*Republic*, 488a-489) who must fend off the efforts of unqualified crewmembers vying for control. This captain-philosopher—the "human being within the human being"[183]—serves as the authoritative ruling part of the tripartite soul, regulating and unifying the lower appetitive and spirited parts so that they can best function for the benefit of the whole. Indeed, this ruling part is said to coordinate the elements of human psychology and of the city, so that, "from having been many things [they become] entirely one, moderate and harmonious." (*Republic* 443d-e)

It is in the *Phaedo*, however, where we encounter Socrates' most direct effort to justify the autonomy of the rational soul. Here Socrates argues against Simmias' popular Pythagorian "Attunement" or "Harmonic" proposal,[184] according to which the activity of the soul is said to supervene on, but not reduce

[182] Hackforth's commentary, p. 72.
[183] *Republic* 589b
[184] The argument takes place in the *Phaedo*, 84c-96a

to,[185] the arrangement of the underlying physical parts of a person, just as the harmonies of a lyre supervene on but do not reduce to the underlying arrangement and behavior of its strings.[186] In opposition to this view, Socrates rhetorically asks if "it is any other part of a man than the soul that governs him, especially if it is a wise one?" (*Phaedo* 94b). Obtaining Simmias' agreement, Socrates argues that, if such is the case, the behavior of the soul cannot correctly be understood as an expression or product of the lower elements. For, after all, the soul "directs all the elements of which it is said to consist, opposing them in almost everything all through life, and exercising every form of control; sometimes by severe and unpleasant methods like those of physical training and medicine, and sometimes by milder ones" (*Phaedo*, 94d). Thus Socrates seems to be making the following argument. For any x and y, where the behavior of y supervenes on the status of x, the direction of causation will flow unilaterally from x to y, and, as such, y cannot bear causal influence on the status of x. Since Socrates believes that the soul can, in fact, "direct" and "oppose" the lower parts, the soul's behavior therefore must *not* supervene on the status of the lower parts.

[185] The non-reductive element is an important component of Simmias' argument, for Simmias preserves the intuition that the soul, unlike the parts upon which it supervenes, is nonetheless "divine".

[186] The central thrust of the argument appears at *Phaedo*, 86c-d: "Well, if the soul is really an adjustment, obviously as soon as the tension of our soul is lowered or increased beyond the proper point, the soul must be destroyed, divine though it is; just like any other attunement, either in music or in any product of the arts and crafts, although in each case the physical remains last considerably longer, until they are burnt up or rot away."

(2) Descartes

Whereas Plato's argument for the autonomy of the rational soul derives from the *causal* relationship that obtains between the rational part of the soul and the lower appetitive parts, Descartes' argument for the "real distinction" of the cogito will chiefly be *epistemic* in nature. Recall that in his famous thought experiment, Descartes rejects as unreal everything he can doubt, assuming that his perception of such objects could be the product of a dream or of an "evil demon" deceiving him. Descartes thus finds reason to deny the reality of the external world and his body—both of which might very well be the product of deception. Yet Descartes finally asserts that he cannot doubt his own thinking "I"; this "I" itself cannot be part of a demon's deception. For, as he puts it, if he, Descartes, is being deceived, "[i]n that case I too undoubtedly exist, if he is deceiving me" (*Meditation* II; AT VII, 25; CSM II, 17). That is to say, some "I" must exist in order for it to be the "subject of deception." Descartes elaborates:

> I saw that while I could pretend that I had no body and that there was no world and no place for me to be in, I could not for all that pretend that I did not exist. I saw on the contrary that from the mere fact that I thought of doubting the truth of other things, it followed quite evidently and certainly that I existed; whereas if I had merely ceased thinking, even if everything else I had every imagined had been true, I should have had no reason to believe that I existed. From this I knew I was a substance whose whole essence or nature is simply to think, and which does not require any place, or depend on any material thing, in order to exist. Accordingly, this 'I'—that is, the soul by which I am what I am—is entirely distinct from the body, and indeed is easier to know than the body. (*Discourse*, Part IV; AT VI, 32-33; CSM I 127).

Notice that in this passage Descartes offers at least three arguments in defense of the distinct status of the "I," the "thinking, non-extending thing" (*res cogitans*) with respect to the "extended thing" of his body (*res extensa*).[187] The first affirms that his "I" and his body are distinct because his thinking "I" is indubitable whereas the existence of his body is something he can doubt. Second, Descartes suggests that the persistence conditions of his body and his cogito are different; whereas the persistence of the "I" depends entirely on thought and is dependent on no material substance, the persistence of the body is dependent on its constitutive material. As such, one could exist without the other, and, as Descartes puts: "Two substances are really distinct when each of them can exist without the other."[188] Yet a third argument is implicit here, that is, the *distinct existence* argument Descartes sets out in Meditation IV[189]. Here Descartes attests that he "knows that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it. Hence the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct, since they are capable of being separated, at least by God" (Meditation IV, ATVII; CSM II

[187] And he offers additional arguments elsewhere, for example, the argument from divisibility, *Sixth Meditation* (AT VII 86-87: CSM II 59).
[188] See Geometrical Exposition in Replies to the Second Objections (AT VII 162; CSM II 114). Descartes says, further: "Strictly speaking, a real distinction exists only between two or more substances; and we can perceive that two substances are really distinct simply from the fact that we can clearly and distinctly understand one apart from the other." *Principles*, Part I, Article 60 (AT VIIIA 28; CSM I 213) Passages found in Hoffman's intriguing discussion in "Descartes's Theory of Distinction," *Philosophy and Phenomenological Research*, Vol. LXIV, No. 1, January 2002, p. 58.
[189] My treatment follows Matthews (1992) and Hoffman (2002).

54). That is, Descartes believes that if he can understand clearly and distinctly *any individual object*, then, since God would not imbue him with a faulty faculty of understanding, God is capable of creating that very object just as Descartes clearly and distinctly understands it. As such, since Descartes believes he can clearly and distinctly understand both the "thinking, non-extending thing" of his "I" and the "extended thing" of his body[190]—that is, insofar as Descartes understands them as two distinct objects—these two objects, are, at least in principle, distinct.

(3) Kant

Now let us briefly re-examine Kant's argument in favor of the autonomy of the self-conscious "I" and notice how it draws upon *both* Plato's causal argument and Descartes' epistemic argument. Recall Kant's assertion, as quoted in the Introduction: "The fact that the human being can have the "I" in his representations raises him infinitely above all other living beings on earth" (2006, p. 15, §127). And, further, recall his contention that "a human being really finds in himself a capacity by which he distinguishes himself from all other things, **even from himself** insofar as he is affected by objects, and that is *reason*" (1997a, p. 57 4:452, bold mine). Finally, recall Kant's assertion that this "I," with the aid of reason, can "of itself, independently of anything empirical, determine the will"

[190] See Sixth Meditation (AT VII78: CSM II 54): "[O]n the one hand I have a clear and distinct idea of myself, in so far as I am simply a thinking, non-extended thing [that is, a mind], and on the other hand I have a distinct idea of body, in so far as this is simply an extended, non-thinking thing. And accordingly, it is certain that I am really distinct from my body, and can exist without it." (AT VII 78: CSM II 54).

(1997b, p. 37, 5:42). Now, inasmuch as Kant's rational "I" can "of itself independently of anything empirical" determine the will, I take him to be subscribing to Platonic causal argument. And insofar as Kant believes a person can distinguish himself *from* himself, i.e., his rational nature from his empirical nature, I take him to be affirming the epistemic positions espoused by Descartes.

(4) Distinct Object Thesis and Contemporary Review

Now, I wish to suggest that the historical line of support recommending the distinct and autonomous status of the reasoning "I" with respect to the facts of the "me," has drawn from one compelling, implicit, underlying assumption. That is, inasmuch as one's body, one's attitudes—or, collectively, one's "facts"— can become for the self-reflecting "I" an *object*, either of the understanding or of agential influence, then such facts must be regarded as *distinct from* the "I". Let us call this assumption the *Distinct Object Thesis*. More formally,

> *(DOT)*: *For any x and any y, if x can become an object of y, where y can observe or act upon x, y must to some degree possess an existence that is distinct from x.*

I will not attempt here to defend the general metaphysical merits of this thesis, as such would exceed the scope of this thesis. But I contend that such a proposition has been implicitly invoked in support of the historical arguments. And I believe that DOT has reverberated through the contemporary literature. Thus we recall Velleman's epistemic assertion that "consciousness just seems to open a gulf between subject and object, even when its object is the subject himself. Consciousness seems to have the structure of vision, **requiring** its object to stand

across from the viewer—to occupy the position of *Gegenstand* [object or thing]" (2008, p. 179, emphasis mine). And Velleman's claim that the agent's involvement is defined in terms of his "interactions with these very states and events, and the agent's interactions with them are such as they **couldn't** have with themselve*s*" (2000, p. 125, emphasis mine). Of course Korsgaard has pushed for the same claim when she affirms that "[o]ur capacity to turn our attention on to our own mental activities is also a capacity to distance ourselves from them." And finally Armstrong has proposed a similar idea, viz., "it is impossible that the introspecting and the thing introspected should be one and the same mental state. A mental state cannot be aware of itself, any more than a man can eat himself up."[191]

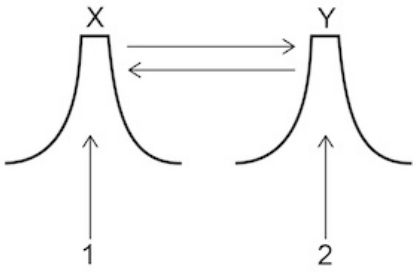
Yet, notice that DOT stands in tension with the case I have been developing over the course of the preceding chapters. For I have argued that the behaviors of the "I" should be understood as extensions of the facts of the "me," and have suggested—though not yet argued—that the "I" itself should be understood as an extension or product of such facts. But notice that if this latter proposition is correct, then it would seem to contradict DOT. To put the point more plainly: If the "I" *can indeed* regard and act upon the "me" as object—either (1) DOT is correct and the "I" and the "me" are in fact distinct, or (2) It is possible for an x to become an object for itself, possible for an x to know and act upon itself—thus contradicting DOT. I will pursue (2). That is, I will argue that while the "I" can indeed observe and act on the "me," the "I" and the "me" should *not* be regarded

[191] Armstrong, D.M. (1968) *A Materialist Theory of Mind* (Routledge: 1968), p. 324.

as existentially distinct—rather, the "I" should be understood as consisting of nothing but the facts of the "me". To clarify and defend this claim, I will first present three models intended to capture candidate relationships between the "I" and the "me"'—*Unilateral*, *Complimentary*, and *Holistic*. I will advocate for the Holistic model, as it best accommodates the proposition that the "I" and the "me" are neither functionally nor existentially distinct. In the following chapter I will apply this Holistic model to a set of metaphysical claims, and, in doing so, attempt to flesh out this model, rendering it more plausible and palatable.

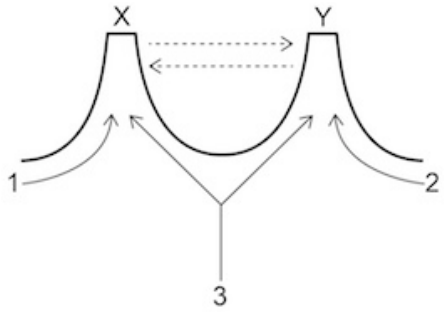
6.3 Three Relations

1. *Unilateral Relation*: Recall from Chapter 2 my characterization of "unilateral" causal relations. I suggested that we understand such a relation as follows: X exercises unilateral causal influence upon y insofar as the causal powers that x bring to bear on y do not derive from y. I illustrated this relationship by way of the behavior of billiard balls. For our present purposes, it will be more helpful to employ the behavior of waves. Thus, imagine two distinct adjacent water bodies, 1 and 2. And imagine that each water body, when agitated, produces a wave—call them x and y (x wave corresponds to water-body 1; y wave corresponds to water-body 2).



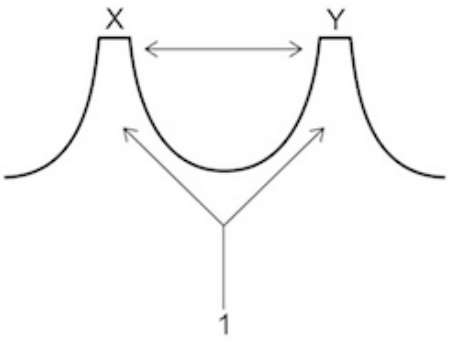
Now, when wave x strikes wave y, x will bring to bear upon y a set of causal properties that are wholly distinct from the causal properties of wave y. That is to say, the causal powers that x bring to bear on y will not derive in any way from y (and vice-versa)—and this because that which gives rise to the causal powers of x (water-body 1) is entirely distinct from that which gives rise to the causal powers of y (water-body 2). Moreover, note that if we were to remove wave y and water body 2 entirely, this would bear no impact on either wave x or water body 1. As such, we may correctly say (1) That the causal influence wave x bears on wave y and vice-versa will qualify as unilateral and (2) That a necessary condition on their unilateral causal influence will be their status as distinct existences.

2. *Complimentary Relation*: Now let us consider a different, complimentary, relation between x and y. Here we will regard x and y as two waves that both *partially* feed off of the same water-body (water-body 3), but also feed off of two distinct water-bodies (water-body 1 and water-body 2).



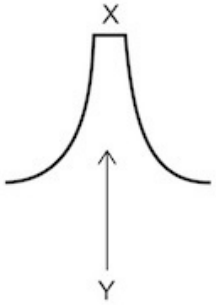
Now, again, notice that both x and y *partially* feed of causal source 3 but that x and y also feed off, respectively, distinct sources 1 and 2. As such, the causal properties that x bring to bear on y (and vice-versa) will derive in part from the common source 3 (should the intensity of 3 increase or diminish, so also would the causal properties of both x and y) and *in part* from their distinct water-bodies. As such, in the event that x and y collide, the causal influence that x will bear on y and vice-versa will be *partially* distinct, or, we might say, *partially unilateral* (hence the dotted line). Furthermore, given that the content and behavior of x and y will be *partially* determined by the same causal source, we may say that x and y lay claim to *partially* distinct existences.

3. *Holistic Relation*: Finally, let us consider the Holistic relation. Again, imagine two waves, x and y. Yet this time imagine each of them drawing on the *same* water body, and, as such, arising as an expression of *the same* causal source, water-body 1.



Now let us consider the causal and existential implications of this model. Note, first, that we may be tempted to speak of two distinct waves, and, indeed, of two

distinct waves bringing distinct causal powers to bear on each other. After all, ***in a relative sense***, such talk would be justified. Wave x might, for example, be much heavier than wave y. Or wave y might be moving much faster than wave x. ***Yet notice that the causal properties of each individual wave derive from and are expressions of the same causal source: water-body 1***. As such, while we might speak of distinct causal properties belonging to each wave, such claims would be spurious. In fact, we cannot properly speak of either wave possessing causal properties that are in any way distinct from water-body 1, or, by association, distinct from each other. To see this, notice that if water-body 1 were suddenly to become still, each wave would collapse; if water body 1 were suddenly to become much more agitated, the agitation would communicate appropriately to x and y. To this extent, *appearances notwithstanding*, neither wave x nor wave y can exert *any* unilateral causal force on the other. And, further, again appearances notwithstanding, neither wave can lay claim to a distinct existence. Rather, wave x, wave y and water-body 1, must be understood as *unitary*, or of-a-piece: they rise and fall *as one*. Indeed, to the extent that these two waves must be understood as expressions of the same corporate entity, we might just as well—and more simply—make use of the following illustration.



Here we will understand wave x as nothing but a *formal aspect* of water-body y.

Having drawn out these three forms of relation, we may now turn our attention back to the "I" and the "me" and ask which model most accurately maps their relationship. Clearly, the conception of the "I" and the "me," as presented by the Rationalist thinkers, is most naturally reflected in the Unilateral or Complimentary models. For each of these models preserves at least *some* of the causal autonomy and existential distinction of the "I". Yet the Holistic model most clearly accommodates the relationship I have been advocating. For on this model the "I" can neither be said to exert any unilateral causal power over the "me," nor to possess any existential distinction with respect to the facts of the "me." Here again, the "I" cannot be understood as something "over and above" the facts; it does not possess a "transcendent" relation to the facts. Rather, it is just an expression, or product, of the facts. Indeed, as Arpaly, following Hume,[192] has put it:

> The tendency to believe that we are captains of our souls has a rival tendency in the human heart—the tendency to believe that we are, as it were, slaves of our souls, that we have fates and identities and inner voices that we cannot escape with our laughable reasoning capacities. […] But we are not the captains of our souls, nor are we servants of our souls. Quite simply we are our souls. (Arpaly, 2003, p. 179)

We "are" our souls, Arplay suggests. That is, what we "are" is *just* the constellation of our facts. And Blackburn, responding to the charge that a Holistic

[192] "I may venture to affirm of the rest of mankind that they are nothing but a bundle or collection of different perceptions, which succeed each other with an inconceivable rapidity, and are in perpetual flux and movement" (*T*:1.4.6, p. 165).

conception of the self leads to an attitude of "passivity,"[193] offers this similar claim:

> We might advance this charge [that we are passive with respect to our desires or our "parts"] because of the grip of the Kantian picture, only now thinking of ourselves, as Captains, suddenly subordinate to a successfully mutinous crew. Or similarly, we might be thinking of ourselves in effect as only the shell within which desires are lodged: the ship which a possibly alien crew is working and directing. But this is wrong: the person is the *totality* composed of body and form, or ship and crew. (1998, p. 251, emphasis mine)

Again, as Blackburn suggests, we are "the totality". That is, while the "I" of self-consciousness may be capable of recognizing the facts of the "me" and even going to work on such facts—this "I" is not itself ontologically distinct from the sum total of the facts. Indeed, even William James, after sketching the distinct capabilities of the "I" and the "me," cautions against regarding these two parts of the self as ontologically distinct. As he puts it: "I call these [the "I" and the "me"] 'discriminated aspects,' and *not separate things*, because the identity of the *I* with me, even in the very act of their discrimination, is perhaps the most ineradicable dictum of common sense." (2001, p. 43, emphasis mine)

"Common sense" for some, perhaps. But for many of us, for a variety of reasons, the "I" *does indeed* feel, and regard itself, as something very much distinct from the facts of the "me". For, again, perhaps nothing seems more self-evident than that "we," that the status of our "I," cannot be reduced to or explained away by an enumeration of our facts. Yet I hope, in the concluding two chapters, to (1) Make a more persuasive metaphysical case for the Holistic model

[193] I will address this "passivity" charge at greater length in the next two chapters.

and (2) Argue that the Holistic model boasts a number of practical and ethical advantages over the alternatives.

CHAPTER 7

METAPHYSICAL IMPLICATIONS

7.1 Introduction

"Where *in* the world is a metaphysical subject to be found?" Wittgenstein asks.[194] I anticipate that a consideration of the Holistic model will inspire this very question. That is, if, as Arpaly and Blackburn have suggested, I *just am* my facts, then how can we understand the "I" of self-reflection, the "I" that James says *is* conscious; the "I" that regards itself as a "single party," and that can observe and attend to my facts? What sense can we make of this very "I," so familiar to experience, that seems so conceptually and phenomenologically incommensurable with facticity? In this chapter I will attempt to shed some light on these problems, exploring some ways in which the Holistic model can reconcile the "I" to the facticity of the "me" and thereby find its "place in the world"; or, to put it a little differently, "find its place *as* the world". I will proceed as follows. First I will revisit Dennett's Holistic "Joycean Machine" and outline the "I"'s role as "virtual captain," suggesting, as he does, that the "I" should *not* be understood as a distinct "single-party" but rather as a Humean bundle or corporation of "facts". With this in mind, I will address a series of concerns associated with "Practical Reason and Temptation," "Self-Integration," "Responsibility and Incontinence" and the

[194] Ludwig Wittgenstein, *Tractatus Logico-Philosophicus*, (trans.) D.F. Pears and McGuiness, B.F., (intro.) Bertrand Russell, (New York: Routledge Classics, 1974). Passage from §5.633. For further treatment of Wittgenstein's exploration of the "I," see Vohra (1986) and Lockhorst (1991).

conceptual problems involved in "Subject-Object" dualism.[195] In discussing such matters, I will try to fill in the Holistic picture, and strengthen my case in support of it.

7.2 The Joycean Machine

The question, again, is: Assuming a person is nothing over and above his facts, how do we make sense of the unitary-seeming "I" which seems so resistant to facticity? To address the question, let us first consider a non-human animal of the Kantian and Korsgaardian variety, a creature which, as Kant says, does not have an "I" in its representations and whose will, as both Kant and Korsgaard suggest, is entirely governed by the drives or impulses of *bruta necessitas*. Let us suppose that a toad, Sam, fits this description. Sam is a biological machine; a product of evolution; a homeostatic collusion of drives and mechanisms—of, collectively, "facts"—aimed at survival and procreation. When Sam is threatened by a predator, he hops away, when Sam is injured he attends to his wound. Indeed,

[195] My account will not attempt to dispel all of the "mysterious features" of the "I". Nor will it attempt to explain the "I" entirely away. My only object will be to offer a non-monolithic characterization of the "I". For similar accounts, see Dennett (1991, esp. p. 288); Damasio (2010); Thomas Metzinger, *Being No One: The Self-Model Theory of Subjectivity*, (Cambridge: MIT Press, 2004), p. 403. And Buddhist accounts, as represented, for example, by Chogyam Trungpa. Consider Trungpa's Hume-friendly illustration: "The experience of oneself relating to other things is actually a momentary discrimination, a fleeting thought. If we generate these fleeting thoughts fast enough, we can create the illusion of continuity and solidity. It is like watching a movie, the individual film frames are played so quickly that they generate the illusion of continual movement. So we build up an idea, a preconception, that the self and other are solid and continuous. And once we have this idea, we manipulate our thoughts to confirm it, and are afraid of any contrary evidence." *The Myth of Freedom and the Way of Meditation*, (ed.) John Baker and Marvin Casper, foreword Pema Chodron (Boston: Shambhala Publications, 1976), p. 13.

Sam is capable of a wide variety of behaviors designed and governed entirely by his biological processes to ensure his survival. Again, per Kant and Korsgaard's understanding, Sam is not self-conscious. As such, while Sam's cognitive apparatus registers representations of the external world—and, as such, permits Sam to interact in manifold ways with the world—his cognitive system does not register self-representations, and, for this reason, Sam cannot know, observe or interact with himself[196]. That is, Sam does not know that "he" possesses particular beliefs or desires; he cannot engage in self-reflection; he is just drawn to or repelled by aspects of the world, as dictated by his biological processes and basic cognitive skills. Recognize, further, that since Sam possesses no "I" that "stands for" or "represents" the complex of his facts. As such, when we refer to Sam, all we are referring to is a set of drives, desires, fears, cognitive mechanisms, etc., none of which involve, or spring from, or attach to, any self-aware "Sam".

Now let us consider a self-conscious human being, a "person". Let us call her Joyce. Like Sam, Joyce's behavior is governed by a network of drives, impulses and cognitive mechanisms—collectively, "facts". But Joyce, the "person," possesses a very substantial mechanism that Sam does not possess. For unlike

[196] Note again that I am just stipulating, for the sake of argument, that Sam the toad does not possess even minimal self-consciousness; I'm not suggesting that toads do or don't actually possess it. For an interesting treatment of the origins and gradual development of sentience and self, see Damasio's discussion of "core" selves and "autobiographical selves" (2010). He says: "I am ready to believe that whenever brains begin to generate primordial feelings—and that could be quite early in evolutionary history—organisms acquire an early form of sentience. From there on, an organized self-process could develop and be added to the mind, thereby providing the beginning of conscious minds. Reptiles are contenders for this distinction, for example; birds make even stronger contenders and mammals get the award and then some" (p. 27). See also, Metzinger (2003) and Baker (1998).

Sam, who *cannot* assume a self-conscious perspective, Julia *can* assume one—that is, Joyce can have an "I" in her representations, or assume the perspective of "Joyce". Not only can Joyce achieve a self-conscious perspective, but, as Kant and Korsgaard observe, Joyce, by means of "Joyce," can know herself, attend to herself and "represent" or "stand for" herself. Let us then understand the process of Joyce becoming "Joyce" as an act of Joyce "taking herself up into herself" and thus becoming both a self-representation and self-representative.[197] Indeed, when Joyce/"Joyce" utters the sentence, "I am Joyce,"[198] she is referring *not just* to a set of facts in dynamic interaction, but to a unitary subject, a personal "I,"—a single emissary or voice—who stands for, speaks for, and regards herself as, to some extent, distinct from her facts. As Damasio puts it: "The notion of a large collective of wills expressed through one single voice is not mere poetic fancy. It connects with the reality of our organisms where that single voice does exist in the form of the self in a conscious brain." (2010, p. 39)

Now, the Rationalists contend that when "Joyce" attends to the facts of Joyce, "Joyce" is exploiting a distance or a "space" between "her" and such facts. She *must* be doing so, so DOT affirms, for when "I" observe or act upon something, "I", the observer and actor, cannot be identical to the object "I" am observing or acting upon. Further, it is in virtue of the putative "distance" separating "Joyce"

[197] See Metzinger here: "For the first time, system-related information now becomes globally available as system-related information, because the organism now has an internal image of itself as a whole, as a distinct entity possessing global features" (2005, p. 4).

[198] Recall again from 2.3: "My active employment of both pronouns *I* and *me/"me"* to designate my objective identity will reflexively implicate the subjective, self-identifying "I". Or, to put the matter a bit differently, my recognition of my objective existence *as me* necessarily implicates my recognition of my subjective-reflective existence as *"I"*."

from her observed facts, that Joyce, operating as "Joyce," can "dominate" or assume *power over* such facts, and, in so doing, "insert" Joyce[199] into the causal order. Yet I have militated against such a characterization. Indeed, I have suggested that, while Joyce, via "Joyce," may very well subject Joyce's facts to analysis, while she may endorse, or identify with, or otherwise "throw her weight behind" her facts, the course and outcome of such activity will redound *not* to the autonomous wherewithal of "Joyce," but, rather, to the total, Holistic coordination of Joyce's corporate facts. As such, "Joyce," as Dennett has claimed, and as I shall go on to elaborate, possesses *no distinct unitary existence*; she is just, as Dennett avers, a "virtual captain of crew," or emissary, that represents, at any moment, the resultant coalition of Joyce's dispositions.[200] That is to say, at any given moment, the self-aware person, or agent, "Joyce," who regards herself as the "leader" or "driver" of Joyce, will just express the resultant global orientation of Joyce's facts—as Damasio has put it, "a collective of wills expressed through one single voice" (2010, p. 39). I grant that this may seem obscure; hopefully it will become less so over the course of the following discussion.

---

[199] Consider a puppeteer manipulating the strings of his puppet. Thus "Joyce" manipulates the strings of Joyce, such as to dictate Joyce's role in the causal order.

[200] See Damasio's elaboration: "The oddest thing about the upper reaches of a consciousness performance is the conspicuous absence of a conductor *before* the performance begins, although, as the performance unfolds, a conductor comes into being. *For all intents and purposes, a conductor is now leading the orchestra, although the performance has created the conductor—the self—not the other way around*. The conductor is cobbled together by feelings and by a narrative brain device…" (2010, p. 25, emphasis mine).

7.3 On Practical Reason and Temptation

Our sense of ourselves as unitary autonomous agents perhaps most saliently presents itself in the context of practical reasoning and in our experience of the "push and pull" of our conflicting attitudes. For it is here where we experience ourselves as *subject to* the various persuasive tendencies of our nature, *burdened with* the task of resolving the tensions between them, and *working* to achieve this end. Indeed, it is perhaps in virtue of our phenomenology of *struggling with*, *overcoming*, or *succumbing to*, our conflicting forces, that we most convincingly regard ourselves as autonomous "operators," intermediating among our facts. How can the Holistic picture make sense of this phenomenology?

Notice first that when we engage in deliberation, when we must "make up our minds" about a given matter, we commonly regard ourselves as monarchic judges—as arbiters—perched "above the fray" and witnessing the behavior of conflicting "attorneys". Thus John's "I,"—from *John's Future*—operating as John (3), wishes to regard himself as occupying a position of impartial Kantian detachment, gazing down at the conflicting testimonies of John (1) and John (2), such that, as Blackburn has put it, "the whole crew is then within [his] purview" (1998 p. 260). Indeed, from such a detached position, John (3) may very well desire to "hear" the appeals of John (1) and John (2) with an impartial, dispassionate, ear, and he may wish to evaluate their appeals on the basis of pure, impartial reason, such that his reason will "of itself, independently of anything empirical, determine the will" (Kant, 1997b, 5:42). Yet, we have seen from Chapter 3, that, in his efforts to engage pure, impartial reason, John (3) will

encounter redoubtable, and often undetectable, opposition. Furthermore, even if John (3) *were* able to engage pure, impartial reason, it remains to be seen how such a calculative faculty, detached from John's "contingent profile of concerns," could enable him to resolve his dilemma. That is, in order for John (3)—the purely rational Kantian Captain—to evaluate the merits of each prospective career choice, he will eventually need to appeal to John's idiosyncratic values, desires, fears, longings, regrets, hopes, etc. As such, the deliberative efficacy of John (3)—Johns "I"—will be heavily contingent on the dispositions of John (1) and John (2). And, this, again, is to be expected. For as Blackburn puts it: "You, when you deliberate, are whatever you are: a person of tangled desires, conflicting attitudes to your parents, inchoate ambitions, preferences and ideals, with an inherited ragbag of attitudes to different actions, situations, and characters. You do not manage, ever, to stand apart from all that." (1998, p. 252)

So, the "I" that pulls back may not be as "independent" or as "unitary" as it seems to us. And yet we are still left with the puzzling question of how to understand our *experience* of "taking sides with," "identifying with," or "throwing our weight behind" a particular desire or principle. After all, when we do such things, it certainly feels as if "I"—distinct from a particular attitude—am *interacting with*, or *operating upon*, such an attitude. How, then, can we make sense of this phenomenology? To address the question, I will make use of another illustration: *Cookie*.

Tom is overweight and suffers from high blood pressure. His doctor has told him that if he doesn't reduce his blood pressure he is going to have to take

medication that can cause “sexual problems”. One alternative means of reducing his blood pressure is through weight-reduction. So Tom has gone on a diet. Yet this hasn’t been easy for Tom. Indeed, after a very frustrating and depressing day at work, Tom enters his home to see his wife pulling a tray of chocolate chip cookies out of the oven. Tom smells the cookies and feels a powerful urge to eat one. Let us call this urge “Desire.” Suppose “Desire ” says: “Eat the cookie.” But just as Tom experiences the tug of desire, he remembers the doctor’s warning, and an opposing force, “Fear” proclaims: “Don’t be an idiot, resist the cookie!” Now, as “Fear” looms up, “Desire” begins to subside. But then Tom catches another whiff of the cookies and “Desire” pipes up: “Oh, c’mon! One cookie won’t hurt! After all, you’ve had a rough day. You deserve a little pleasure! It will help you relax. And a little relaxation is also important.” So now “Desire” begins to wax and “Fear” to wane. Then Fear responds: “You know if you eat one, you will be very tempted to eat two. C’mon! Just stick to your resolution!” But then “Desire” kicks in again: “Aw, c’mon, you were good all day, cut yourself a little slack.” Then suddenly Tom’s wife calls out, impatiently: “Oh c’mon Tom! What? Are you just gonna stand there, staring at the cookies? Just have one, it won’t kill you!” Alarmed by his wife’s “Shaming” complaint, and encouraged by her endorsement of “Desire” (for she is aware of his diet) Tom decides that he might as well just eat one cookie. So he “throws his weight behind” “Desire” and takes one.

Such, in any case, is our commonsense, rationalist, account of Tom’s behavior. Again, in short: “Reasonable Tom” has engaged in a painful battle

between two conflicting forces vying for his allegiance. "Tom" listens to "Fear," and considers its appeals; then "Tom" listens to "Desire" and considers its appeals. The appeals of each disposition are equally strong and they exert an equal force on "Tom". Finally, Tom's wife's voice pipes in, making him feel foolish, which prompts "Shame." With the support of "Shame" and his wife's expressed endorsement of "Desire," "Tom" decides to "throw his weight" behind "Desire" and thereby overcomes "Fear."

However, the Holistic model provides a different description. Here too, the global organism, Tom, is indeed, undergoing a struggle because elements of Tom are in conflict. And "Tom," the representative of Tom, registers and endures this conflict. But note, crucially, that the forces in conflict—"Desire" and "Fear" and eventually "Shame"—aren't merely heteronymous forces *within* global Tom: They are actually parts of Tom, speaking **for** him and thus speaking **as** Tom and, to the extent that "Tom" recognizes such voices, speaking s "Tom". As Shariff, Schooler and Vohs put it:

> Instead of saying that my consciousness (me) is making the decisions, we need to say that I am conscious of the parts of my brain (still me) that are making the decisions. Instead of saying the "I" moves the machinery of my brain, "I" *am* the machinery of my brain, and "I" consequently move myself.[201]

That is to say, it is not merely "Desire" or "Fear" that looms up, but "Tom *qua* Desire" or "Tom *qua* Fear" that loom up, and, in doing so, assume *global Tom's will*. But notice that, insofar as self-conscious "Tom" gives way to Tom qua Desire or Tom qua Fear, they will be speaking for both "Tom" and Tom (because

[201] Shariff, A. F., Schooler, J., & Vohs, K. D. 'The Hazards of Claiming to Have Solved the Hard Problem of Free Will.' In *Are We Free?: Psychology and Free Will* (Oxford University Press, 2008), p. 93

"Tom" is the representative of Tom). At this point, neither "Tom *qua* Desire" nor "Tom *qua* Fear" has proven sufficiently powerful to completely command Tom's overall will and push him to action. So now Tom sends the Korsgaardian or Frankfurtian "Tom *qua* Impartial Mediator" into the fray, hoping that he will "resolve the conflict" and therefore restore the unity of Tom's will.[202] But how, and to what extent, does "Tom *qua* Impartial Mediator" actually accomplish this? Notice, first, that, as discussed in previous examples, "Tom *qua* Impartial Mediator" does not act as "parade marshal" for Tom's mental proceedings. That is to say, "Tom *qua* Impartial Mediator" neither determines the appearance nor the strength of the attitudes of "Tom *qua* Fear" and "Tom *qua* Desire"—nor, crucially, does he determine either the appearance, or the strength, of his "own" orientation to "Desire" or "Fear". True, "Tom qua Impartial Mediator" may be able to engage a sharper or more concentrated apparatus of reason. He may be able to summon a greater data set and perform various inferences and projections. But, crucially, the extent to which "Tom *qua* Impartial Mediator" can do this *is not*, *ultimately*, *up to* "Tom *qua* Impartial Mediator". For, again—and I suggest that this is the ***critical*** *move*—the extent to which "Tom *qua* Impartial Mediator" can deliberate independently of "Fear" and "Desire" will depend *entirely* on both the strengths of "Tom *qua* Fear" and "Tom *qua* Desire" and on whatever cognitive resources are available to "Tom *qua* Impartial Mediator" (or, for that matter, to Tom *qua* global human being). **As such, "Tom's" contribution will**

[202] Here I am on board both with Korsgaard's assertion that persons do indeed strive for volitional unity, and with Frankfurt's notion that "[i]t is a necessary truth about us that we wholeheartedly desire to be wholehearted." "The Faintest Passion," from (1998), p. 116.

depend entirely on the composite facts of Tom. Or as I put it in Section 5.2, "Tom" operating as the "aristocratic" self that attempts to "rule" over his parts, will *just constitute another part of* the Holistic, or corporate totality of Tom, and the role it plays in "pulling his parts together" will be no less contingent than the roles played by the other parts. To put it a little differently: "Tom" is *always* a subset or expression of Tom; Tom is *never* a subset or expression of "Tom". Indeed, as Shariff, Schooler, and Vohs have put it: "In the former [rationalist] position, one's "I" is understood to refer to "the me that does thing thinking" and this self is credited with being the one that consciously controls one's actions. The new approach dissolves the conscious self into the larger "I" and "the me that does the thinking" is embedded within the whole brain. "I" still control my actions, but the "I" is reconceived to be the coalition of my brain processes." (2008, p.93)

Yet we have still left unexplained the question of how to account for Tom's *feeling* of "throwing his weight behind" "Desire". This is a tricky problem, but I believe the Holistic model is up to the challenge. Note, again, that Tom *qua* global human being is invested in and represented by "Fear," "Desire," and "Shame". That is to say, "Fear," "Desire" and "Shame" both ***represent*** Tom and ***are*** Tom. Yet recall that while Tom and "Tom" are invested in and represented by such contrary factions, Tom/"Tom" desires to be unified and wholehearted.[203] I suggest, thus, that "Tom's" feeling of "throwing his weight behind" "Desire" should just be understood as the relief Tom/"Tom" feels at the decisive resolution

[203] We thus can make sense of Tom accurately expressing the view that he is "of two minds" or even "of three minds" about a given matter.

of the conflict. That is, Tom/"Tom"'s global system, previously in conflict or tension, has ultimately and happily coalesced around the unity-restoring, volition-assuming disposition of "Desire." Now, at least for the time being, as Tom takes his cookie and enjoys it, his disposition will be harmoniously concentrated. Only afterwards, as "Desire" reasserts itself—"Oh c'mon, just one more!"—will "Fear" kick in again and the whole embarrassing, grinding, tearing-yearning-striving dynamic will resume.

Yet one might still wonder, as Socrates does in the *Phaedo*—how to account for the possibility of a person *overriding* his desires or drives, or, indeed, overriding his physical body. As Socrates asks, in such cases, doesn't the soul or the "I" "[direct] all the elements of which it is said to consist, opposing them in almost everything all through life, and exercising every form of control; sometimes by severe and unpleasant methods like those of physical training and medicine, and sometimes by milder ones" (*Phaedo*, 94d)? Yet again, I suggest the Holistic account is up to the challenge. That is, without addressing the difficult question of "downward causation"—how, that is, a mental attitude can causally interact with a subvening physical substance—all the Holistic model needs to demonstrate is that the "I" does not bring *autonomous* or *unilateral* powers to bear on the facts of the "me," physical or otherwise. And this it can do. For, again, we need only understand such behavior as a product or expression of our totality of psychophysical facts—robust inclinations, volitional necessities, intellectual character and what have you. Someone, for example, who suffers from tendencies toward self-hatred and masochism will more likely develop

certain eating disorders; someone who is unusually ambitions will tend to "push himself" harder than others and to "punish themselves" more severely when they fail; a person who suffers from chronic depression will be more inclined, after sustaining a personal trauma, to jump off a bridge. Viewed in this light, I suggest that such behaviors should be understood as perfectly consistent with one's overall economy of personal facts—perfectly "nested," that is to say, in the functional holism of a person's global psychophysical system. Indeed, as I have repeatedly suggested, I believe we should properly understand such attempts to alter or modify such facts as further expressions or realizations of one's facts.

7.4 Self-Integration and Incontinence

The Holistic view will also illuminate questions concerning the nature of identity, and, by association, questions involving incontinence or weakness of will. Recall first that on the prevailing Rationalist view the self is both constructed and held together by the "I" that can "have a hand in" such processes. That is, recall James' claim that the "I" can "appropriate" elements of the "me," and Frankfurt's and Korsgaard's assertions that we can identify with or externalize certain desires or tendencies, establishing "intrapsychic constraints and boundaries," and "pulling our parts together". Again, as I have suggested, there is something to this view, for, unlike un-self-conscious animals, we certainly *can* form an opinion of ourselves and go to work on ourselves. Yet I suggest that the Holistic model of self will much better capture what *actually goes on* in such efforts of self-integration, and, indeed, will better capture our experience.

To begin with, as discussed earlier, note that my very decision to identify with or externalize certain desires and attributes, and my success in doing so, will critically be determined by a set of tendencies over which I will not be able to exercise control. The homophobic man may very well wish to externalize or alienate his desire for other men, but whether or not he can do so will turn on the strength of his underlying desires. The woman who wishes to identify more with her "love" for her father instead of her "anger" toward him may or may not succeed, as determined by the actual strength of her actual feelings. As such, one's identifications and endorsements will still *contribute* to the overall composure of one's identity. But the nature and extent of their contribution will need to be accurately assessed with respect to the overall composite, that is, the strengths, or weights of *all* the facts, whether or not one is conscious of them. That is, on the Holistic model, the workings of underlying and unacknowledged tendencies in our nature should be accorded just as much "authority" with respect to deciding our true identity. As Schechtman has suggested, one's robust inclinations "are to be given presumptive authority even when we do not identify with them" (Schechtman, 2004, 426).

Indeed, this "reorientation" with respect to one's identity will help make sense of our failure to live up to our self-conceptions and will thus make sense of the phenomena of incontinence. To see this, note that the thrust of this problem of incontinence or *akrasia*, as originally introduced by Socrates in the *Protagoras*, lay in the presumption that the knowledge of "our good" should exercise incontrovertible authority with respect to our behavior. As Socrates puts it: "No

one who either knows or believes that there is another possible course of action, better than the one he is following, will ever continue on his present course" (*Protagoras* 358b-c). And he says, further: "[T]he power of appearance makes us wander and often change up and down with respect to the same things and change our minds in our actions and choices." But knowledge, on the other hand, "renders the appearance ineffective and makes our soul remain in the truth" (*Protagoras* 356c-e). As such, once we conclude, or "really know," what is in our best interest, we cannot do otherwise; or, to put the idea differently, if we were to act in a way that was not in our best interest, we must *not* have known what was actually in our best interest.[204] Take again, Cookie. Given the ease with which Tom's appetite tempted him, we must conclude, per Socrates, that Tom must have not "known," that the best course of action was to abstain from eating it. He might, as Davidson[205] postulates, possess *prima facie* belief that it was a bad thing to do, or even an "all things considered belief," but not "all out knowledge," which would have pre-empted even the slightest temptation.

But, again, note that the pull of this problem lay in our endorsement of the presumptive authority of our self-conscious rational conviction. Once, however, we reject this assumption, we can recognize additional, equally authoritative sources, and hence equally decisive sources of motivation.[206] Again, if I may

[204] See supporting accounts in Terry Penner's "Knowledge vs. True Belief in the Socratic Psychology of Action," APEIRON, 1996, 29 (3), p. 199. And R. M., Hare, *The Language of Morals*, (Oxford: Clarendon Press, 1952).
[205] See Davidson "How Is Weakness of the Will Possible?" in Davidson, *Essays on Actions and Events*, (Oxford: Clarendon Press, 1980), pp. 21-42.
[206] See Arpaly: "A theory of rationality should not assume that there is something special about an agent's best judgment. An agent's best judgment is just another belief." From "On Acting Rationally Against One's Better Judgment," *Ethics*,

borrow an episode from personal experience: As a twelve-year-old, fresh from watching, and being terrified by, the movie "Jaws," our family went on our summer vacation to Lake Tahoe. We had gone several summers in a row and I had swum every summer. But that summer I was afraid to enter the water. My parents and all the grownups I knew assured me that "sharks are not freshwater fish". And I believed what they told me. Indeed, I was just as sure of the truth of this proposition as I was that I had ten fingers and ten toes. Indeed, I would have placed a bet on the fact—would have wagered an entire month's allowance on it. And yet, in spite of my knowledge, I could not as much as dip a toe in the lake. Here Socrates might contest that, as a matter of fact, I must not have "really known" that sharks could not swim in this water. But this seems absurd, and, in any case, such an assertion would seem to trivialize Socrates' claim; that is, if *by definition*, one cannot act against what one knows, then no argument can be adduced otherwise.

Indeed, according to such a view, we can also make sense of Arpaly's analysis of Huck Finn's "inverse akrasia". That is, in spite of Huck's thoroughly considered, rational endorsement of his desire to turn Jim in to his owner; in spite of his self-identification as someone who respects the rights of slave-owners, and regards slaves as property—in spite of this, he ultimately yields to his desires or "wills" which were a source of shame and which he wished to externalize or

---

110: 488-513, p. 512. See also Stocker: "Motivation and evaluation do not stand in a simple and direct relation to each other, as so often supposed." And, further, "their interrelations are mediated by large arrays of complex psychic structures, such as mood, energy, and interest." Michael Stocker, "Desiring the Bad: An Essay in Moral Psychology," *Journal of Philosophy*, 1979, 76: 738-753.

alienate. Recall that, on Korsgaard's account, to violate your consciously endorsed principles "is to lose your integrity and so your identity, and to no longer be who you are. That is, it is to no longer be able to think of yourself under the description under which you value yourself and find your life to be worth living and your actions to be worth undertaking. It is to be for all practical purposes dead or worse than dead" (1999: 101). Yet, again, once we have unseated the sovereignty of the rational "I," and allowed other parts equal authoritative claim to our identity—parts that, as Schechtman recommends, include even our unconsciously-held beliefs and desires—we can appreciate the possibility that Huck's behavior was not at all a violation of his "true" nature, but a faithful expression and affirmation of it.

7.5 Responsibility

The above considerations lead naturally to a discussion of responsibility, that is, the question of whether or not a person is genuinely or "ultimately" credit or blame-worthy for their actions. Recall in Chapter 4 that I argued that persons cannot "produce" or "author" their attitudes or actions such as to merit ultimate responsibility; and I argued that compatibilist accounts of "assuming ownership" failed to head off undermining worries about one's formative causes—that is, various conditions that give rise to the development of one's reasons-response mechanism or one's character. As such, in both cases, I argued in favor of the proposition that one should *not* be regarded as ultimately responsible—ultimately blame or credit-worthy—for one's attitudes or actions. Again, I recognize this

may not "sit well" with our intuitions. That is, given our powerful phenomenology of *doing things*, given our sense that *we determine* what we do, and given our standard—and perhaps irresistible—custom of blame and praise attribution,[207] we may naturally resist such a proposal. Yet I hope that a further discussion of the Holistic model will render it more plausible and palatable.

First, I acknowledge that a robust self-consciousness renders human persons capable of an enormous range of behaviors that are inaccessible to non-self conscious creatures. For example, creatures lacking self-consciousness cannot possibly desire to "control their drinking," "deal with their anger," or "manage their weight". Moreover, self-conscious creatures will possess an incomparably greater degree of understanding with respect to the consequences of their actions; for example, they will be able to project how their actions will affect their relationships, their social standing, etc. Indeed, self-consciousness will even be required for a conceptual grasp of such practices as punctuality, promise-keeping, and forgiveness. As such, insofar as human beings are capable of, say, entering into agreements or showing up on time, they can and should be held accountable for failing to do so. That said, one can hold another "accountable" for an action without holding the other "responsible" for it. To help bring out the difference, I will make use of an illustration, *Julia and Jane*.

Julia and Jane are friends and colleagues at the same real estate agency. On a number of occasions, when Julia and Jane have arranged a time to meet socially,

[207] See: Peter Strawson's "Freedom and Resentment" in Gary Watson, (ed.), *Free Will* (Oxford: Oxford University Press).

Jane has arrived on time but Julia has arrived between ten and fifteen minutes late. At first Jane dismisses Julia's tardiness, attributing it to traffic conditions, or some other circumstance, but she finds herself growing impatient. Finally, Jane confronts Julia, informing her of the fact that she has grown frustrated with her tardiness. In response, Julia offers a sincere apology. "I'm sorry," she says, "It's just that I'm a chronically late person. It's a problem I have and I've been working really hard on it." Moved by Julia's sincerity—if a bit perplexed by her explanation—Jane accepts Julia's apology. Yet the next time they meet, Julia again shows up late, and again offers the same apology: "I'm so sorry. It's a real problem I have and I've been working on it. It's just really hard for me to be on time for anything." Again, Julia comes across as genuinely contrite, but this time Jane is less moved by her apology. After all, it occurs to Jane that Julia almost always arrives at work on time, and, to her recollection, Julia has never complained of missing a flight. Finally, after Julia shows up late once again, Jane asks for a rundown of Julia's activities during the hour preceding their meeting. "Well," Julia responds, "I promise you, I had every intention of coming on time. In fact, I even set aside an extra fifteen minutes. But just as I was about to leave, I got a call from my friend who was having a panic attack. I was going to talk to her on my cell on the way, but my cell was out of power and I'd misplaced my charging cord and…" As Julia speaks, Jane suddenly realizes that at least part of the reason Julia is late is that Julia simply doesn't value their friendship enough to do what is necessary to arrive on time (after all, again, Julia arrives to work on time and never misses a flight). This upsets Jane, and it occurs to her that perhaps

she should threaten not to meet Julia socially anymore if she can't guarantee punctuality. On the other hand, given that Jane enjoys Julia's company, and given that they need to sustain a good working relationship, Jane decides against this. In any case, it occurs to Jane that she cannot justifiably blame Julia for her actions; her behavior is a natural expression of her character and priorities, and these are things that "Julia" cannot determine. At this point Jane may consider expressing her regret that Julia does not sufficiently value their relationship, or else trying to do something to further strengthen their friendship so as to seem more deserving of respect, but Jane realizes that neither of these gestures may prove sufficient, indeed, that they may backfire. After all, Jane has already made her wishes plain, and they have already been friends for years, and it seems unlikely Jane can do anything to render her friendship more important to Julia. As such, with a heavy heart, Jane accepts Julia for who she is, and their friendship for what it is, and from then on simply expects Julia to show up late to their meetings and plans accordingly.

So, yes, individuals can recognize right and wrong, and recognize the consequences of their actions. Moreover, individuals are capable of modifying their behavior such as to accommodate their understanding of such norms. For these reasons, persons can and should be held accountable for their behavior. The motorist who has driven drunk, lost control of his car, and struck a pedestrian, has indeed behaved irresponsibly and recklessly, and must be held to account. He should be fined; his license should be suspended or revoked; perhaps he should serve jail time. The sex offender has indeed committed crimes and must be

brought to justice, sent to an institution and kept away from young children. Similarly, those who have performed honorably, or who achieve great success, should be praised and esteeme. After all, their actions have stemmed from *them* and have expressed *their agency*. And yet, as Strawson, Pereboom, Smilansky and others have argued, for a person to be ultimately responsible for his or her actions—for them to be *ultimately* blame or credit-worthy—a person will need to be seen as *responsible for the extent of the agency* he or she can exercise. And this, as I have argued, is beyond human reach.

7.6 Subject and Object - The Conceptual Problem

I have tried to demonstrate how we may understand both the activity and the identity of the "I" as an expression of the facts of the "me." Yet nothing I have said so far has directly addressed the problem of our phenomenology of "observing" and "acting upon" our facts; for we *certainly do* experience ourselves as "looking" at ourselves and "acting" upon ourselves. Following Descartes here, I suggest that this phenomenology is unmistakable and unshakable: "I"—however "I" am to be conceived—am, *in fact*, observing. And in the act of observing, as DOT affirms, it seems conceptually impossible that the observer (whether unitary or pluralistic) can be identical to that which is observed. Here, I will try to shed some light on this problem, but I regret that I will not be able to shine very much light.

Consider, first, that, consistent with the conceptual claim raised above, the "I," *per* subject, cannot know itself *as such*. That is, insofar as the "I" observes, *that*

which it observes can only stand before it as *object*; hence, the "I" cannot, as Hume says "catch itself" in the act of observation.[208] Wittgenstein has put this point nicely. Comparing the observing "I" to a physical eye, he says: "You will say that this is exactly like the case of the eye and the visual field. But really you do *not* see the eye. And nothing in the visual field allows you to infer that it is seen by the eye" (1974, §5.633, p. 69, emphasis his). That is, while the "eye" can view objects in the world and even view itself as object, it cannot see itself "seeing". As such, Wittgenstein suggests that "[t]he subject does not belong to the world: rather it is a limit of the world" (1974, §5.632, p. 69). And since, on the *Tractarian* view, "[t]he world is the totality of facts,"[209] we therefore cannot speak of the existence of the "I" and, as such, it "must be passed over in silence" (1974, §7, p. 89). Note, also, that I have argued against the unitary conception of the "I" and pushed instead for a view of the "I" as a contingent bundle or "coalition" of facts. As such, insofar as we can allow for the existence of an observing "I," I suggest again that we should only recognize it as a transient coalition or composite. To such an extent, either (1) We may just have to remain "silent" on the metaphysical "I" or subject, or (2) The perceiving "I" should be recognized as synonymous with the facts of the "me".

Unfortunately, neither of these proposals substantially alleviates the central worry. Indeed, no matter how much we may wish to dismiss or dispel the

[208] See Hume (*T*: 1.4.6, p. 165): "For my part, when I enter most intimately into what I call myself, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch myself at any time without a perception, and never can observe any thing but the perception."

[209] Wittgenstein (1974, §1.1, p. 5)

existence of the "I," we *simply cannot escape* our phenomenology. Nothing seems more self-evident than the fact that "I," the perceiving subject, *am indeed perceiving*; that "I" *am indeed feeling*; that "I" *am indeed thinking*. Indeed, as Kant puts it, "The **I think** must **be able** to accompany all my representations; for otherwise something would be represented in me that could not be thought at all, which is as much as to say that the representation would either be impossible or else at least would be nothing for me" (1998, B132, p. 146, emphasis his). Furthermore, even supposing that the "I" can be re-described as a "coalition of facts," we must still account for the possibility of a "coalition of facts" assuming a unique perspective *from which it can observe* features of the "me". Earlier, I nodded at some clumsy explanations, such as the self "taking itself up into itself," and nodded to Metzinger's proposal that: "For the first time, system-related information now becomes globally available as system-related information, because the organism now has an internal image of itself as a whole, as a distinct entity possessing global features" (Metzinger, 2005, p. 4). But I recognize that such explanations hardly illuminate. Indeed, I take consolation in the fact that Schopenhauer regarded this problem as an inextricable, "knot of existence," and "miracle par excellence."[210] And so I will leave it at that.

[210] He writes: "Now the identity of the subject of willing with that of knowing by virtue whereof (and indeed necessarily) the word "I" includes and indicates both, is the knot of the world (Weltknoten), and hence inexplicable. For to us only the relations between objects are intelligible; but of these two can be one only insofar as they are parts of a whole. Here, on the other hand, where we are speaking of the subject, the rules for the knowing of objects no longer apply, and an actual identity of the knower with what is known as willing and hence of the subject with the object, is *immediately given*. But whoever really grasps the inexplicable nature of this identity, will with me call it the miracle 'par excellence.'" Schopenhauer (1974, p. 211)

CHAPTER 8

ETHICAL IMPLICATIONS

8.1 Introduction

Over the course of the preceding chapters I have challenged the Rationalist conception of the self, working to undermine the impartial, rational, unitary "I". In its place I have offered a Holistic model of the self, a model that assigns no privileged, directorial authority to the "I" and indeed, understands the "I", and its reasoned preferences, as a transitory, mercurial expression of one's global coalition of empirical facts. In fleshing out this Holistic model, I have tried to show how it can accommodate various puzzles associated with deliberation, temptation, identity, weakness of will and responsibility. Now, in this final chapter, I will spell out some of practical and ethical advantages of the Holistic model.

Let us begin with this passage by Simon Blackburn, as it nicely captures what I regard as the key ethical and practical failing of the Rationalist (or Kantian) picture:

> If we see our fellow human beings as each possessed of Kantian control, and only succumbing to other pressures when things are going wrong, then a dangerously optimistic politics is possible. The implication is that because our fellows are fundamentally able to guide themselves by rational restraint, then of course they *ought* to be safe with guns, or drugs, or motor cars, or sacrosanct areas of private behavior. They have themselves all that is required for self-control and reason. The unhappily common failures, when people shoot each other, abuse drugs, drive unsafely, or brutalize their families show us only defectives who unaccountably will not listen

> to the voice of reason within them, and these can safely be demonized, put away, rejected as beyond the social pale. We thus combine unreasonable optimism about what people might be like, with unreasonable hatred of them when they are not like that. (1998, p. 268)

Notice how dramatically Blackburn's claims contrast with Korsgaard's injunction that "[a]s a rational being, as a rational agent, you are faced with the task of *making something of yourself*, *and you must regard yourself as a success or a failure insofar as you succeed or fail at this task*" (2009, p. xii, emphasis mine). Indeed, in light of Korsgaard's passage, we may begin to appreciate Blackburn's warning that the Kantian/Rationalist model may give rise to a "dangerously optimistic politics." In this passage, Blackburn gestures at some of the consequences of such dangerous optimism, suggesting, for example, that it is unrealistic for us to expect of persons that they should exercise the optimal degree of rational self-control, and that, as such, it is unfair for us to "demonize" and "reject" such persons after they have failed to do so. In the following sections, I will further explore the dangers of such unrealistic expectations and argue that the Holistic model provides an approach that, while perhaps not as robustly optimistic, is nonetheless more practical, accommodative, and compassionate.

8.2 Unrealistic Expectations

"There's a touch of the divine in being an agent,"[211] Korsgaard opines, and I believe her comment captures a powerful, familiar, and jealously guarded,

[211] Korsgaard (2009, p. 85). And see Chisholm who asserts that free agents possess a quality "which some would attribute only to God: each of us, when we act, is a prime mover unmoved." From "Freedom and Action," in K. Lehrer (ed.), *Freedom and Determinism* (New York: Random House, 1966), p. 11-44.

intuition. That is, our immediate, pre-reflective, sensation of *being*, does not *feel* restricted, does not *feel* empirically, or even perhaps temporally,[212] bound. We feel, in a word, "stringless". Moreover, insofar as we take ourselves to be essentially unconstrained and capable of directly commanding our actions, we likewise regard ourselves as unbound Sartrean agents, fee to "make ourselves" as we choose. From an early age, our elders reinforce this premise, assuring us that nothing necessarily stands in the way of the achievement of our dreams. Yes, we may run up against some resistance, and we may have to put in some hard work, but as long as we are willing to bear down—and as long as we receive a few lucky breaks along the way——there is nothing, at least no *internal* impediment, to our progress.[213] Of course there is some truth buried in these assertions. We *do* set our sights on our objectives, hard work *will* in most cases help us achieve our ends, we *can* overcome a great deal of resistance, and our confidence in our boundless potential *can* serve as powerful motivation. Yet I suggest that such an outlook, in spite of its optimism, will bear several associated costs—costs, indeed, that, in many respects, will significantly outweigh the benefits.

Note, first—as I have argued in various forms over the course of this thesis—that, for empirical human beings composed such as we are, our possibilities are, in fact, quite radically constrained. Yes, hard work will in most cases help us progress toward the realization of our goals, yet, even under the very best of

[212] See Dennett's nice description of our being "moral levitators". Daniel Dennett, *Freedom Evolves* (Penguin, 2004). And see Schopenhauer (1969, Vol. 1, p. 280) on one's experience of occupying the *nunc stans*.

[213] See, for example, Malcolm Gladwell's popularization of the "10,000 Hour Rule," the idea that 10,000 hours of "deliberate practice" should allow any otherwise capable person to become an expert in a given field. Malcolm Gladwell, *Outliers, The Story of Success* (Back Bay Books, 2011).

conditions, given our native capabilities, such hard work will only pay off to a particular extent. Anyone who has struggled to grasp a mathematical principle, or to find the "right word" to complete a poem, or to perform the "moonwalk" dance move, or even to ask a girl or boy out on a date, will recognize that all human beings are not born with the same basket of talents. This, again, is not to suggest that such difficulties cannot be overcome. But the question of *whether* one can overcome them, and how *quickly*, and by what *means*, will greatly depend on a set of antecedent factors and on an array of external conditions such as our upbringing, our familial, educational, financial and political support structures. Indeed, once we recognize that what we achieve is, as such, only in a very relative sense, and only to a certain degree, "up to us," we can begin to establish more realistic, practical, and healthy expectations. If, for example, a student recognizes that he struggles with writing, but excels in mathematics, he can make a better informed judgment with respect to where to devote his energies (he can, for example, ease up on his study of mathematics, and devote more time to working on his writing; or else he may decide to really exploit his mathematical talent and let his English skills languish). If one doesn't feel attracted to the opposite sex but to one's own, and if one can embrace this fact, one won't need to waste energy fending off "alien desires" or chastising oneself for feeling them, but can begin to engage in more natural and fulfilling relationships.[214] If one is sick, and one can't muster a "positive attitude" toward one's condition, one needn't feel guilty about

[214] Recall Schechtman: (2004, p.425-426) "[T]he work of shaping a life is less of a task of micro-management. It is less about directly settling conflicts than about establishing safe boundaries within which these conflicts can be allowed to play themselves out."

it. As such, the capacity to *acknowledge* and *accept* certain core tendencies or "robust inclinations," can greatly serve to reduce one's internal conflict and to thereby render possible a more wholehearted, passionate life.

While such acts of recognition and acceptance will help establish conditions for a healthier relationship to oneself, they can also dramatically serve to improve one's relationships with others. Consider, for example, the common habit of "just ignoring" or else "trying to fix" unattractive qualities in a prospective romantic partner. In the first case, especially after the early romantic glow has worn off, those undesirable qualities (annoying habits, social quirks, etc.) one has tried to deny or push out of view likely will erupt into vexing prominence. In the latter case, one may dedicate an enormous amount of energy to amending a lover's set of foibles—usually with very minor success. "Fixer-uppers" often turn into personal "breaker-downers," and, in any case, the assumption of such a critical, corrective approach to one's lover may well serve to undermine his or her self-esteem, and, as such, undermine the good faith of the relationship. Consider, second, how such an approach will beneficially apply to child raising. That is, from a very early age, children clearly demonstrate marked differences in character, talents, preferences, temperament, etc. One child may manifest athletic skills and an interest in team sports, another may demonstrate more intellectual skills and an interest in academics; one child may behave in a more domineering or bossy manner, another in a more submissive or accommodative manner. As such, I suggest that a "one approach fits all" policy in child rearing will prove inferior to an approach tailored to each child's peculiarities. Of course there is no

harm in exposing all children to a standard spectrum of activities, but expecting that all children respond similarly is unreasonable, just as is "forcing" a child to enjoy or to perform well at something for which he feels no inclination and shows minimal talent. Indeed, forcing a child to persevere at some such activity may ultimately serve to poison the activity, predisposing him or her against it for years to come. Finally, a deep appreciation of other peoples' natures can render us more capable of accepting and adjusting to their offenses. That is to say, the more we can understand the mechanisms that give rise to a person's behavior—childhood trauma, dispositions of temperament, etc.—the easier it may[215] become for us to regard their offenses as symptomatic and general in nature, and not take them as personally: The crazy neighbor lashes out at us, not because there is anything particularly offensive about us, but because she is crazy and because we happen to live next-door.

8.3 Some Objections and Replies

My discussion of the benefits of such "realistic expectations" may well prompt the following multi-part objection: "Your Holistic assessment of human nature with all its constraints is bleak and self-defeating. I mean, just consider the following: (1) Suppose I realize that I actually don't care for my character traits. In that case, your Holistic model wouldn't offer much in the way of hope. Even worse, your view will license a complacent, or self-defeating attitude. I mean, why bother doing anything at all to change or improve myself, if, as you say, I am

[215] I say it "may" help us view people as such, but not necessarily. Consistent with the Holistic view, the extent of our understanding may or may not allow us to overcome or modify our reactive attitudes.

unable to do so, since such self-transformative change isn't actually in the cards? Further, (2) If your view is correct, then no one is ultimately responsible for who they are or how they act, and that just can't be right, because of course we are responsible for the kind of people we are, and for our actions. At least I know I am! And even worse, that kind of thinking—that is, not holding people responsible for who they are and how they act—is not only wrong, it is dangerous. I mean, how else are we supposed to enforce laws if no one is "responsible" for breaking them? (3) If your account is correct, if we are "just our facts," then what renders us superior, I mean, more worthy of dignity and respect, than non-human animals? For certainly, as Kant has said before, human beings *really are* in possession of a unique dignity and are uniquely deserving of respect, such that we have an obligation to treat persons not as a means, but as ends in themselves. And, finally, (4) No matter how persuasive your arguments may be, your account just can't be correct. It can't be correct because I *just know* that I'm more than a bunch of facts—physical and psychological or otherwise. I mean, there's a *real me in here*! And this real me is—in some way I admit I can't really explain—truly and extraordinarily special.

I will address these concerns in order. I will call them: (1) The Objection from Complacency or Self-Defeatism, (2) The Objection from Responsibility, (3) The Objection from Dignity and Respect, and (4) The Objection from Specialness.

8.3.1 The Objection from Complacency or Self-Defeatism

The Objection from Complacency or Self-Defeatism expresses the following worry. If all I am is my "facts," and if even my efforts to change my facts will be conditioned by more facts—then to a great degree,[216] whatever I can become has been predetermined. I mean, I may appreciate some formal transformation—just like, for example, a caterpillar turns into a butterfly, or a tadpole turns into a frog—but even these modifications, however dramatic, will just consist in the general unfolding of my facts, like a photograph's details emerging on a sheet of paper. And this can seem awfully demoralizing to one who doesn't approve of one's facts. Indeed, to embrace such a view would seem to support an attitude of defeatism: "If I can't change myself, then why bother doing anything at all?"[217] Indeed, such conditions might even be invoked to justify unethical behavior, viz., if "nothing is controllable," then "everything is permitted."[218]

But the Holistic model does *not* license such complacency or pessimism, and this for a number of reasons. First, recall from Chapter 3 that we rarely, if ever, possess a sufficient grasp of our facts. That is, we neither possess an accurate account of our present facts, nor an accurate account of any additional facts that may emerge in the future. After all, it is not uncommon for people to realize late

[216] I provide this qualification to account for indeterministic or "random" factors.
[217] See empirical research by, for example, Kathleen D. Vohs and Jonathan W. Schooler, that purport to show that accepting determinism encourages cheating. "Encouraging Belief in Determinism Increases Cheating," (*Psychological Science*, 2008) Vol. 19, Number 1, p. 49-54.
[218] See A. F., Schariff, J. Schooler, Kathleen D. Vohs. (2008). "The Hazards of Claiming to Have Solved the Hard Problem of Free Will," in *Are We Free?: Psychology and Free Will* (Oxford: Oxford University Press), p. 182.

in life that they possess certain passions or talents they never suspected they possessed or could ever possess. Thus we find people who eventually discover that they possess a love for the arts, or, indeed, love for members of the same gender. Take the following episode from my own life. By the age of 40, due to a terrible history with dogs, I had developed the conviction that I was "not a dog person," that I even "lacked the capacity" to feel anything but aversion in the presence of a canine. And yet, due to the overwhelming demand of my wife and kids, on condition that I would not be expected to take on any of the walking or feeding duties, I eventually, begrudgingly, agreed to adopt a Springer Spaniel puppy. Over time, in spite of my most stubborn efforts to "hold on" to my aversion, and "fend off" any feelings to the contrary, the creature eventually won me over, toppling one supporting, if minor, column of my self-conception. As such, insofar as we recognize our incomplete grasp of our facts, and recognize that latent or unexpressed facts may yet emerge, we need not yield to complacency or pessimism.

Notice, further, that, even if one were to correctly recognize all of one's facts, and recognize that one will never succeed at overcoming one's undesirable attributes, this very understanding will not necessarily *render* one complacent or despairing. For, if among our set of facts are tendencies to stubbornness, optimism and idealistic thinking, these very tendencies may well continue to assert themselves, overpowering our tendencies toward depression or pessimism. As Nagel puts it: "What sustains us, in belief as in action, is not reason or justification, but something more basic than these—for we go on in the same way

even after we are convinced that the reasons have given out."[219] Thus, for example, we find people who may fully recognize that they have experienced no success with dieting—who, indeed, have chronically followed the same hapless trajectory of hope, weight loss, lapse and weight gain—but, who, *in spite of this knowledge*, start all over again with the same zealous, imperturbable optimism, certain that "This time will be different!" Or else consider people who cycle through the same dysfunctional romantic relationships—people who can spell out in extraordinary detail their self-destructive patterns of behavior—and yet who, again and again, embark on new relationships utterly convinced that "This time it will be different!" As such, if a disposition of hope and optimism is "built into" someone's character—no matter how unjustified they, themselves, regard such hope and optimism—they need not worry about losing it.[220] Indeed, recall the powerful influence of Timothy Wilson's "psychological immune system" (2002, p. 38) that protects us from threats to our psychological wellbeing, whose central rule is: "Select, interpret and evaluate information in ways that make me feel good." (2002, p. 39)

8.3.2 The Objection from Responsibility

I have argued in defense of the proposition that to possess "ultimate" responsibility for an attitude or an action requires of one that he can *produce* or *own* the attitude or act in question in such a way that it redounds not to "forces at

[219] Nagel, from "The Absurd," from *Mortal Questions*, (Cambridge: Cambridge University Press, 1979), p. 20.
[220] I grant that this may generate a certain degree of cognitive dissonance, but even such dissonance will likely fade beneath an overpowering tendency toward optimism and idealistic thinking.

work" in a person, but to the person "himself". Of course, per the Holistic view, persons can do no such thing. And yet, as I discussed in the previous chapter, the fact that one should not be regarded as "ultimately responsible" for one's actions does not preclude us from holding people "accountable," and requiring that people "answer for" their actions. That is, the motorist who knows that drunk driving is dangerous and illegal but who has nonetheless driven drunk, lost control of his car, and struck a pedestrian, has indeed behaved irresponsibly and recklessly, and must be held to account. The sex offender must be brought to justice, committed to an institution, or otherwise kept away from young children. Indeed, in such cases, perpetrators of such crimes are certainly exercising agency; their actions are properly attributable *to them*. And yet, for a person to be deeply or ultimately blameworthy or ultimately creditworthy for his or her actions, he or she will need to be seen as *responsible for the extent of the agency he or she has exercised*, and this, for the reasons I have enumerated, exceeds an individual's powers.

As such, while we must hold such people to account, and bring them to justice, I suggest that such justice should not be administered according to "retributive" or "punitive" principles, but rather according to "restorative" principles, with the intent of rehabilitating the wrongdoer and establishing conditions whereby he or she can mend relationships with his victim and community.[221] Of course, justice

[221] Note that practices of restorative justice have shown a greater rate of success than those of retributive justice. See, for example, Lawrence W. Sherman and Heather Strang, *Restorative Justice: The Evidence* (Smith Institute: 2007). See also Howard Zehr *Changing Lenses: Restorative Justice For Our Times*, (Herald Press; 25th Anniversary Edition, 2015)

administered in this way may not relieve our vindictive desire to inflict harm and suffering upon the wrongdoer—to impose, as Nietzsche puts it, the "metaphysics of the hangman"[222]. But once we have taken a sufficiently wide step back and properly viewed a person in light of antecedent conditions, we should recognize that such vindictiveness is unwarranted.[223] As such, while, again, the Holistic picture denies "ultimate" responsibility—and associated blame and credit—it has nothing to say against accountability, the necessity and legitimacy of law-enforcement, and the capacity for rehabilitation.

8.3.3 The Objection from Respect and Dignity

A further objection to the Holistic model is that it seems to license a dangerously reductive and disrespectful view of human beings. We may understand such a threat coming from two directions. First, by unseating the autonomous "I," the Holistic model can threaten our confidence in the Kantian proposition that all persons possess a unique dignity and are thus worthy of unconditional respect. In the second place, one may contend that the Holistic

---

[222] Nietzsche, *Twilight of the Id*ols*/The Anti-Christ* (trans.) R.J. Hollingdale, (intro.) Michael Tanner (Penguin, 1990), from "The Four Great Errors," §7, pp. 58-70.

[223] Spinoza is wonderful here: "I have laboured carefully, not to mock, lament, or execrate human actions, but to understand them; and, to this end, I have looked upon passions, such as love, hatred, anger, envy, ambition, pity, and the other perturbations of the mind, not in the light of vices of human nature, but as properties, just as pertinent to it, as are heat, cold, storm, thunder, and the like to the nature of the atmosphere, which phenomena, though inconvenient, are yet necessary, and have fixed causes." Baruch Spinoza, *Tractatus Politicus*, (ed. and intro.) R.H.M. Elwes, (trans.) A.H. Gosset (London: G. Bell &Son, 1883), Chp. 1, §4.

model can give rise to prejudicial thinking, i.e., the idea that people from a certain country, or belonging to a certain gender, or who are of a certain skin-color, predictably present a definite set of character traits.[224] I will address these concerns in order.

Let us first flesh out the Kantian conception of universal respect for persons. Recall Kant's assertion, quoted in the Introduction, that "[t]he fact that the human being can have the "I" in his representations [i.e., is self-conscious] raises him infinitely above all other living beings on earth. Because of this he is a *person* […] i.e., through rank and dignity an entirely different being from *things*, such as irrational animals…" Indeed, Kant more directly spells out his conditions for respect in the *Groundwor*k, as follows:

> Now morality is the condition under which alone a rational being can be an end in itself, since only through this is it possible to be a lawgiving member of the kingdom of ends. Hence morality, and humanity insofar as it is capable of morality, is that which alone has dignity… for nothing can have a worth other than that which the law determines for it. But the lawgiving itself, which determines all worth, must for that very reason have a dignity, that is, an unconditional, incomparable worth; and the word respect alone provides a becoming expression for the estimate of it that a rational being must give. *Autonomy* is therefore the ground of the dignity of human nature and of every rational nature. [4:435-6]

According to Kant, thus, irrational creatures are mere expressions of brute drives—they possess no knowledge of, and cannot legislate over, such impulses. "Worth" is something that can only be created by creatures that are capable of legislating over their impulses; indeed, the very *act* of legislation itself *creates* such worth, and, as such, the *capacity* to legislate—that is, autonomy—is the

[224] We observe this kind of thinking on particularly noxious display in Nietzsche's discussion of lambs and birds of prey in his *Genealogy of Morality* (2006). See §13 pp. 25-28.

precondition of all worth. Since Kant believes that human beings are uniquely capable of autonomy, they themselves are the only "worthy" creatures, and, as such, they alone are in possession of a dignity that is "beyond all price"[225] and are deserving of unconditional respect.

Yet I have argued at length that our exercise of "autonomy" or "self-legislation" is not as Kant—or Rationalists in general—characterize it. That is, we certainly can "step back," and "legislate" over our impulses, but this act of legislation (1) Never proceeds free from our empirical nature; and, as such, (2) Is never entirely "up to" an autonomous "I" that operates free of heteronymous influences. That being the case, we *need not* relinquish our faith in the proposition that all persons are in possession of dignity and are worthy of respect. For, after all, why should we regard *autonomy* as the sole dignity and respect-conferring virtue? Why can't we ground it elsewhere?[226] For example, why not ground respect in the capacity for sentience? After all, it seems quite natural for us to respect the dignity and rights of human beings who have *not* developed a robust self-consciousness and who have not developed the capacity to self-legislate—or, indeed, who have lost such capacities—and we do so in part because we do not wish to see such human beings suffer. Further, it seems quite natural to respect the rights, the interests and feelings, of non-human animals, to preserve their natural habitats, to advocate against animal cruelty and the horrific conditions of factory

[225] See Kant, (*Groundwork*: 434:32, p. 42)

[226] Damasio is nice here: "Perhaps the most indispensable thing we can do as human beings, every day of our lives, is remind ourselves and others of our complexity, fragility, finiteness, and uniqueness. And this is of course the difficult job, is it not: to move the spirit from its nowhere pedestal to a somewhere place, while preserving its dignity and importance; to recognize its humble origin and vulnerability, yet still call upon its guidance." (1994, p. 252)

farms.[227] As such, we may locate the wellsprings of our respect not in our appreciation for law-giving autonomy, but in our desire for the non-suffering of all sentient creatures.[228]

With respect to the second concern about stereotyping and prejudice, it might seem that Kant's proposal would be especially useful in the service of forestalling such tendencies. For, again, according to the Kantian view—and very much in keeping with Sartre—no individual, let alone a group of individuals, can be judged exclusively on the basis of facts. We are all, in essence, *noumenal* beings, free of empirical taint, and capable of unfettered self-legislation. Yet I don't think we need to reach so far. First, empirical evidence, and common sense, suggests that stereotypes are overwhelmingly just that—prejudicial, groundless, generalizations. Second, we may assume as a matter of epistemic humility that we are never justified in prejudging an individual on the basis of a social stereotype or otherwise. Indeed, given how ignorant we are of our *very own natures*, we likewise should assume that we possess an incomplete understanding of others. Finally, once we have predicated respect on something as general as sentience, hopefully we can see through our differences, and focus instead on our shared concerns.

---

[227] The Sioux tribe accords an equal respect, not only to non-human animals, but to all "living" creatures, the set of which includes water and stones. See: John (Fire) Lame Deer and Richard Erdoes, *Lame Deer, Seeker of Visions* (Pocket Books, 1994).

[228] See Schopenhauer's marvelous defense of such a "compassion-based" foundation of ethics and his critique of Kantian deontology in Arthur Schopenhauer, *On the Basis of Morality*, (trans.) E.F.J. Payne, (intro.) David E. Cartwright (Cambridge: Hackett Publishing Company, 1995).

8.3.4 The Objection from Specialness

Finally, there just seems to be *something*—something *in* me, or rather, something that *is me* that must reside over and above my facts, something that possesses an extraordinary, irreducible, perhaps spiritual value, that cannot simply be explained away or reduced to facticity. I touched on this concern in my treatment of the "conceptual problem" of subject and object in Section 7.6. In this concluding section, I will spell out a few more proposals that seem to safeguard the irreducibly special character of the "I".

Consider, first, as Nagel does, the certain "primitive amazement" he feels at recognizing the fact that "the universe should have come to contain a being with the unique property of being me" (1986, p. 56). How is it, Nagel appears to be asking, that, over the fathomless reaches of space and time, some combination of insensate matter gave rise to a *personal revelation of being*, and not just *any* personal revelation, but *his* personal revelation. Indeed, Nagel goes on to argue, his personal revelation of being seems incommensurate with any objective description of the world, any description, that is, that consists in an exhaustive inventory of objective, impersonal facts. That is: no descriptive list of impersonal facts seems to add up to *him*, to *his* subjective experience, and, further, any description of the world that would omit a description of his subjective experience would be incomplete.[229] As such, one's personal experience seems to be an ineliminable part of the world.

[229] See also Baker: "If I attribute first-person reference to myself, my sentence cannot be adequately paraphrased by any sentence that fails to attribute first-person reference to me. The attribution of first person reference to oneself seems

A related puzzle touches on the previous conceptual and phenomenological issues raised earlier. That is, if I am composed of nothing *but* the world—if all that I am, and all that I perceive, should just be understood as further attributes of the world—then how can I make sense of the experience of *my* observation of the world; that is, my observation of the world as an entity that is *distinct from me*? How, that is, should the locutions: "I move *through* the world," "I *look out* at the world," "I try to *fit into* the world," or even "I *know* the world," be understood? Again, when we spoke of ocean waves, we referred to the wave "passing over" the ocean, but this description was really just a matter of linguistic convenience, as we understood the wave as only an expression or "aspect" of the ocean. As such, if what I am is just an expression or "aspect" of the world, we would expect such self-affirming descriptions to be similarly specious. Yet, insofar as I cannot help but regard myself as something that is at least in part *distinct from* the world; as something that *perceives* the world, *walks* through the world, etc., I cannot overcome this sense of division; my very consciousness seems to affirm my distinct ontological status with respect to the world.[230] Indeed, perhaps it is this

---

to be ineliminable" (1998, p. 331). And see Metzinger's ponderous response (2003, p. 370): "One way of paraphrasing this sentence according to the current model would be: 'This system currently uses a sentence in a public language to refer to a certain capacity, namely, the capacity to access the content of certain opaque, cognitive simulations integrated into an already existing transparent self-model by higher-order cognitive operations, which then in turn can be integrated into this self-model.'"

[230] Camus puts the point nicely: "If I were a tree among trees, a cat among animals, this life would have a meaning, or rather this problem would not arise, for I should belong to this world. I should **be this world** to which I am now opposed by my whole consciousness and my whole insistence upon familiarity. This ridiculous reason is what sets me in opposition to all creation." Albert

inescapable sensation of our distinction that undergirds the "burden of our freedom" that Sartre, Kant, Korsgaard and others speak of—the fact that, try as we might, we cannot slough off our sense of distinct "being," cannot just dismiss our need to "act" and "assume responsibility" for our actions. However we contort ourselves, we cannot escape from these activities or "hand them off" to impersonal forces working "on" or "through" us. As a result, our unique existence may well impress us as an inescapable burden that we must "heave" through time.

Yet, we may nonetheless regard ourselves as, as Camus tells us, "stronger than [our] rock."[231] Indeed, as the Existential thinkers have said, this burden of consciousness, this inescapable, non-transferable burden of our distinct *being*, may be regarded as a gift. Thus, Pascal writes: "Man is only a reed, the weakest in nature, but he is a thinking reed. There is no need for the whole universe to take up arms to crush him: a vapor, a drop of water is enough to kill him. But even if the universe were to crush him, man would still be nobler than his slayer, because he knows that he is dying and the advantage the universe has over him. The universe knows none of this."[232] Man "knows," Pascal reminds us, and, in virtue of this knowledge, we can pursue projects, nurture relationships, and achieve goals that are entirely inaccessible to creatures lacking in self-consciousness. No other creature—as far as we know—can produce works of art, can practice

---

Camus, *The Myth of Sisyphus and Other Essays*, (trans.) Justin O'Brien (New York: Vintage, 1955), p. 51, emphasis mine.

[231] Albert Camus, (1955), p. 112.

[232] Blaise Pascal, *Pensees* (trans. and intro.) A. J. Krailsheimer (Penguin: 1966), p. 95, §200.

philosophy, can marvel at the wonder of his own transient existence in a dazzling universe. True, our self-consciousness may not permit us to become the masterful, bounded, autonomous creatures we may wish to become, but it still allows us to behave as persons in a personal world, that, for all of its limitations and darkness, is also full of astonishing depth and beauty.

BIBLIOGRAPHY

Alston, William. (1989) "The Deontological Conception of Epistemic Justification" in *Essays in the Theory of Knowledge*, 115-52. (Ithaca: Cornell University Press).

Anscombe, G.E.M. (1976) *Intention*, 2d ed., Ithaca: Cornell University Press.

Aristotle. (2002) *Nicomachean Ethics* (trans.) Christopher Rowe, and Sarah Broadie (Oxford: Oxford University Press).

Armstrong, D.M. (1968) *A Materialist Theory of Mind* (Routledge).

Arpaly, Nomy. (2000) "On Acting Rationally Against One's Better Judgment," *Ethics*, 110: 488-513.
-------, (2003) *Unprincipled Virtue: An Inquiry into Moral Agency* (Oxford: Oxford University Press).

Atkins, Kim and Mackenzie, Catriona, (eds.). (2008) *Practical Identity and Narrative Agency* (Routledge).

Baker, Lynne Rudder. (1981) "Why Computers Can't Act," *American Philosophical Quarterly*, Vol. 18: 157-163.
-------, (1998) "The first-person perspective: A test for naturalism." *American Philosophical Quarterly* 3: 327–346.
-------, (2006) "Moral Responsibility Without Libertarianism," *NOUS* 40:307-330.
-------, (2011) "First Personal Aspects of Agency," *Metaphilosophy*, Vol. 42, Nos. 1-2, January.
-------, (2013) *Naturalism And The First-Person Perspective* (Oxford: Oxford University Press).

Blackburn, Simon. (1998) *Ruling Passions* (Oxford: Oxford University Press).

Boyle, Matthew. (2009) "Active Belief," from *Believe and Agency* (ed.) David Hunter, *The Canadian Journal of Philosop*hy.

Broadie, S. and Rowe, C. (trans.). (2002) *Aristotle Nicomachean Ethics: Translation, Introduction and Commentary* (Oxford: Oxford University Press).

Buckareff, Andrei A. (2004), "Acceptance and Deciding to Believe," *Journal of Philosophical Research*, Volume, 29.

Buss, Sarah, and Overton, Lee (eds). (2002) *Contours of Agency: Essays on Themes of Harry Frankfurt* (Cambridge: MIT Press).

Camus, Albert. (1955) *The Myth of Sisyphus and Other Essays*, (trans.) Justin O'Brien (New York: Vintage).

Cassam, Quassim. (2014) *Self-Knowledge for Humans* (Oxford: Oxford University Press).

Chisholm, Roderick. (1996) "Freedom and Action" from K. Lehrer (ed.), *Freedom and Determinism* (New York: Random House.

Cottingham, John. (2002) "Descartes and the Voluntariness of Belief" *The Monist*, 85:3.

Damasio, Antonio. (1994) *Descartes' Error, Emotion, Reasoning and The Human Brain* (New York: Avon Books, 1994).
-------, (2010) *Self Comes to Mind* (New York: Random House).

Davidson, D. (1980) *Essays on Actions and Events* (Oxford: Clarendon Press).

Demos, Katherine E. (et. al). (2012) "Individual Differences in Nucleus Accumbens Activity to Food and Sexual Images Predict Weight Gain and Sexual Behavior," *The Journal of Neuroscience*, April 18, 2012, 32(16): 5549 –5552.

Dennett, Daniel C. (1985) *Elbow Room, The Varieties of Free Will Worth Wanting* (Cambridge: Cambridge University Press).
-------, (1991) *Consciousness Explained*, (New York: Little, Brown).
-------, (2004) *Freedom Evolves* (Penguin).

Descartes, Rene. (1970) *Descartes: Philosophical Letters*, (trans.) Anthony Kenny (Oxford: Clarendon Press).
-------, (1985) *The Philosophical Writings of Descartes Vol. II* (trans.) John Cottingham, Robert Stoothoff, Dugald Murdoch, (Cambridge: Cambridge University Press).

Dutton, Donald G. and Aron, Arthur P. (1974) "Some Evidence for Heightened Sexual attraction Under Conditions of High Anxiety." *Journal of Personality and Social Psychology* 30, no. 4: 510-17.

Duval, T.S. and Wicklund, R. A. (1972) *A Theory of Objective Self-Awareness* (New York: Academic Press).

Edwards, Paul. (1961) "Hard and Soft Determinism" from Hook, S. (ed.) *Determinism and Freedom in the Age of Modern Science* (New York: Collier).

Feldman, Richard. (2001) "Voluntary Belief and Epistemic Evaluation." In *Knowledge,Truth, and Duty*, (ed.) Matthias Steup, 77-92. (Oxford: Oxford University Press)

Festinger, Leon. (1957) *A Theory of Cognitive Dissonance,* (Stanford University Press).

Fischer, John Martin and Mark Ravizza. (1998) *Responsibility and Control: A Theory of Moral Responsibility*, (Cambridge: Cambridge University Press).

-------, (2002) "Replies," *Philosophy and Phenomenological Research*, Vol. 61, No. 2.

Fodor, Jerry. (1983) *The Modularity of Mind: An Essay on Faculty Psychology*, (Cambridge, MA; MIT Press).

Frankfurt, Harry. (1988) *The Importance of What We Care About* (Cambridge: Cambridge University Press).

-------, (1999) *Necessity, Volition and Love* (Cambridge: Cambridge University Press).

-------, (2004) *The Reasons of Love* Princeton (Princeton: Princeton University Press).

Freud, Sigmund. (1920) *Dream Psychology: Psychoanalysis For Beginners*, (trans.) M.D. Eder, (intro.) Andrew Tridon (New York: The James A. McCann Company).

Gertler, Brie. (2011) *Self-Knowledge* (Routledge).

Ginet, Carl. (2001) "Deciding to Believe." In *Knowledge, Truth, and Duty*, (ed.) Matthias Steup (Oxford: Oxford University Press).

Gladwell, Malcom. (2011) *Outliers, The Story of Success* (Back Bay Books).

Goldman, Alvin. (1993) *Readings In Philosophy & Cognitive Science* (MIT Press).

Gopnik, Alison. (1993) "How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality" from (ed.) Alvin Goldman, *Readings In Philosophy & Cognitive Science* (MIT Press), pp. 315-346.

Guignon, Charles. (2004) *On Being Authentic* (Routledge).

Hare, R. M. (1952) *The Language of Morals* (Oxford: Clarendon Press).

Harvey, John H. and Weary, Gifford. (1985) *Attribution: Basic Issues and Applications* (Academic Press).

Haybron, Daniel M. (2007) "Do We Know How Happy We Are? On Some Limits of Affective Introspection and Recall." *Nous* 41 (3): 394-438.

Hieronymi, Pamela. (2006) "Controlling Attitudes." *Pacific Philosophical Quarterly* 87, 45-74.
-------, (2009) "Believing at Will," from *Belief and Agency*, David Hunter, (ed.), *The Canadian Journal of Philosophy* Supplementary Volume 35 p. 153.

Hoffman, Paul. (2002) "Descartes's Theory of Distinction," *Philosophy and Phenomenological Research*, Vol. LXIV, No. 1, January.

Hook, S. (ed.). (1961) *Determinism and Freedom in the Age of Modern Science*. (New York: Collier).

Hume, David. (1981) *Enquiry Concerning Human Understanding*, (ed.) Eric Steinberg, (Indianapolis: Hackett Publishing Co).
-------, (2000) *A Treatise of Human Nature*, (eds.) David Fate Norton and Mary J. Norton (Oxford: Oxford University Press).

Hunter, David, (ed.). (2009) *Belief and Agency* (Canadian Journal of Philosophy: University of Calgary Press).
Affective Introspection and Recall," *NOUS* 41:3.

Hyman, John and Steward, Helen (eds). (2004) *Agency and Autonomy* (Cambridge: Cambridge University Press).

James, William. (1918) *Principles of Psychology, Vol. 1* (Dover Publications: New York).
-------, (2001) *Psychology, The Briefer Course* (Dover Publications: New York).
-------, (2010) *The Will To Believe and Other Essays in Popular Philosophy and Human Immortality* (Digireads.com).

Kant, Immanuel. (1996) *The Metaphysics of Morals*, (trans. & ed.) Mary Gregor, (intro.) Roger J. Sullivan (Cambridge: Cambridge University Press).
-------, (1997a) *Groundwork of the Metaphysics of Morals*, ed. Mary Gregor, (intro.) Christine M. Korsgaard (Cambridge: Cambridge University Press).
-------, (1997b) *Critique of Practical Reason*, (ed.) Mary Gregor, (intro.) Andrews Reath, (Cambridge: Cambridge University Press).
-------, (1997c) *Lectures on Ethics*, (ed.) Peter Heath and J.B. Scheewind, (trans.) Peter Heath (Cambridge: Cambridge University Press).

-------, (1998). *Critique of Pure Reason*, (trans.) Paul Guyer and W. Allan Wood (Cambridge: Cambridge University Press).
-------, (2006) *Anthropology from a Pragmatic Point of View*, (ed.) Robert Loudon (Cambridge: Cambridge University Press).

Kornblith, Hilary. (1998) "What is it Like to Be Me?" *Australasian Journal of Philosophy*, 76: 1, 48-60.
-------, (1999) "Distrusting Reason" *Midwest Studies in Philosophy*, XXIII p. 181-196.
-------, (2012) *On Reflection* (Oxford: Oxford University Press).

Korsgaard, Christine. (1996a) *Creating the Kingdom of Ends* (Cambridge: Cambridge
University Press).
-------, (1996b) *Sources of Normativity* (Cambridge: Cambridge University Press).
-------, (1997) "The Normativity of Instrumental Reason," in Garrett Cullity and Berys Gaut (eds), *Ethics and Practical Reason* (Oxford: Clarendon Press), pp. 215-254
-------, (1999) "Self-Constitution in the Ethics of Plato and Kant," *The Journal of Ethics* 3: 1-29.
-------, (2003) "Philosophy in America at the Turn of the Century: Realism and Constructivism in Twentieth Century Moral Philosophy," *Journal of Philosophical Research*, pp. 99-122.
-------, (2008) *The Constitution of Agency, Essays on Practical Reason and Moral Psychology* (Oxford: Oxford University Press).
-------, (2009) *Self-Constitution: Agency, Identity, and Integrity* (Oxford: Oxford University Press).

Lame Deer, John (Fire) and Erdoes, Richard. (1994) *Lame Deer, Seeker of Visions* (Pocket Books).

Leiter, Brian. (2007) "Nietzsche's Theory of the Will" *Philosopher's Imprint* Volume 7, No. 7.

Liu, JeeLoo and Perry, John, (eds.). (2012) *Consciousness and the Self, New Essays* (Cambridge University Press).

Locke, John. (1979) *An Essay Concerning Human Understanding*, (ed.) P. Nidditch (Oxford: Clarendon Press).

Lockhorst, G.J. C.. (1991) "Wittgenstein On the Structure of the Soul: A New Interpretation of Tractatus 5.5421," *Philosophical Investigations* vol.14 p.14-20.

Long, Todd R. (2004) "Moderate Reasons-Responsiveness, Moral Responsibility, and Manipulation" in Freedom and Determinism, (ed.) Joe Keim-Campbell, Michael O'Rourke, and Savie Shier (MIT Press).

Lyubormirsky, Sonja, Caldwell, Nicole D. and Nolen-Hoeksema, Susan. (1998) "Effects of Ruminative and Distracting Responses to Depressed Mood on Retrieval of Autobiographical Memories," *Journal of Personality and Social Psychology*, Vol. 75. No. 1, pp. 166-177.

Mackenzie, Catriona. (2007), "Bare Personhood? Velleman on Selfhood," *Philosophical Explorations*, Vol. 10, No. 3, September 2007.

Marino, Gordon. (2004) *Basic Writings of Existentialism* (ed. and intro. Marino) (Modern Library).

Matthews, Gareth B. (1992) *Thought's Ego In Augustine and Descartes* (Cornell University Press).

McGonigal, Kelly Ph.D. (2012) *The Willpower Instinct* (New York: Penguin).
McKenna, Michael. (2000) "Assessing Reasons-Responsive Compatibilism," *International Journal of Philosophical Studies*, Vol. 8, No.1: 89-114.

Mele, Alfred. (2000) "Reactive Attitudes, Reactivity, and Omissions," *Philosophy and Phenomenological Research* 61: 447-452.

Metzinger, Thomas. (2003) "Phenomenal transparency and cognitive self-reference," *Phenomenology and the Cognitive Sciences* 2:353-393.
-------, (2004) *Being No One: The Self-Model Theory of Subjectivity* (Cambridge: MIT Press).
-------, (2005) "Precis: Being No One," *Psyche*.

Mill, John Stuart. (1895) *On Liberty* (London: W. Parker and Son, West Strand).

Moran, Richard. (2001) *Authority and Estrangement* (Princeton: Princeton University Press).

Nagel, Thomas. (1979) *Mortal Questions*, (Cambridge: Cambridge University Press).
-------, (1986) *The View From Nowhere*, (Oxford: Oxford University Press).
-------, (1997) *The Last Word* (Oxford: Oxford University Press).

Nietzsche, Friedrich. (1997) *Daybreak*, (eds). Maudemarie Clark and Brian Leiter, (trans.) Hollingdale, R.J., (Cambridge: Cambridge University Press).

-------, (2001) *The Gay Science*, (ed.) Bernard Bernard, (trans.) Josefine Nauckhoff and Adrian Del Caro (Cambridge: Cambridge University Press).
-------, (2002) *Beyond Good and Evil*, (eds.) Rolf-Peter Horstmann, and Judith Norman, (trans.) Judith Norman (Cambridge: Cambridge University Press).
-------, (2006) *The Genealogy of Morality* (ed.) Keith Ansell-Pearson, (trans.) Carol Diethe (Cambridge: Cambridge University Press).

Nisbett, Richard E. and Wilson, Timothy DeCamp. (1977) "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review*, May, Vol. 84, No. 3.

Nussbaum, Martha. (20016) *Anger and Forgiveness: Resentment, Generosity, Justice* (Oxford: Oxford University Press).

O'Connor, Timothy. (1995) *Agents, Causes, Events* (ed.) T. O'Connor (New York: Oxford University Press).

Pascal, Blaise. *Pensees* (1966) (trans. and intro.) A. J. Krailsheimer (Penguin).

Penner, Terry. (1996) "Knowledge vs True Belief in the Socratic Psychology of Action," APEIRON, 29 (3), pp. 199-230.

Pereboom, Derk. (2001) *Living Without Free Will* (Cambridge: Cambridge University Press).

Plato. (2001) *Selected Dialogues of Plato*, (trans.) Benjamin Jowett (Modern Library Classics).

Pronin, Emily, Lin, Daniel Y. and Ross, Lee. (2002) "The Bias Blind Spot: Perceptions of Bias in Self Versus Others," *Personality and Social Psychology Bulletin*, March, Vol. 28, No. 3, pp. 369-381.

Robins, Anthony. (2007) *Awakening the Giant Within* (Free Press).

Rousseau, Jean Jacques. (2004) *The Confession of Jean Jacques Rousseau,* Vol. 1 (London: Privately Printed for Members of the Aldus Society).

Ryan, S. (2003). "Doxastic Compatibilism and the Ethics of Belief," *Philosophical Studies*, 114, 47-79.

St. Augustine. (1871) *City of God*, Vol. 1 (trans.) Rev. Marcus Dodds, M.A., Edinburgh: T.T. Clark, 38, George St. 1871, reproduced on Project Gutenberg: http://www.gutenberg.org/files/45304/45304-h/45304-h.htm#Page_436.

Sartre, Jean Paul. (1957) "Existentialism" excerpted from *Existentialism and Human Emotions*, (Philosophical Library Inc.,), as reprinted in Marino, 2004.
-------, (1993) *Being and Nothingness*, (trans.) Hazel Barnes (Washington Square Press).

Schechtman, Marya. (1996) *The Constitution of Selves* (Cornell University Press).
-------, (2004) "Self-Expression and Self-Control," *Ratio*, Vol. 17, Issue 4.

Schopenhauer, Arthur. (1974) *On the Fourfold Root of the Principle of Sufficient Reason* (trans.) E.F.J. Payne, (Lasalle, Ill: Open Court).
-------, (1995) *On the Basis of Morality*, trans. E.F.J. Payne, (intro.) David E. Cartwright (Cambridge: Hackett Publishing Company).
-------, (2002) *Parerga and Paralipomena* (trans.) E.F.J. Payne (Clarendon Press: Oxford)
-------, (2004) *Prize Essay on the Freedom of the Will,* (ed.) Gunter Zoller, (trans.) Eric F. J. Payne (Cambridge: Cambridge University Press).

Schwitzgebel, Eric. (2008) "The Unreliability of Naïve Introspection," *Philosophical Review*, Vol. 117, No. 2.

Scott-Kakures, Dion. (1994) "On Belief and the Captivity of the Will," *Philosophy and Phenomenological Research*, 54: 77-103.

Shah, Nishi. (October 2003) "How Truth Governs Belief," *The Philosophical Review*, Vol. 112, No. 4.
-------, Shah and Velleman, J. David. (October 2005) "Doxastic Deliberation." *The Philosophical Review*, Vol. 114, no 4.

Shariff, A. F., Schooler, J., & Vohs, K. D. (2008) "The Hazards of Claiming to Have Solved the Hard Problem of Free Will," from *Are We Free?: Psychology and Free Will* (Oxford University Press).

Sherman, Lawrence W. and Strang, Heather. (2007) "Restorative Justice: The Evidence" (Smith Institute).

Silvia, Paul J. and Gendolla, Guido H. E. (2001) "On Introspection and Self-Perception: Does Self-Focused Attention Enable Accurate Self-Knowledge?" *Review of General Psychology*, Vol. 5, No. 3.

Smilansky, Saul. (2003) "Compatibilism: The Argument from Shallowness," *Philosophical Studies* 1115 (3): 257-82.

Smith, Michael. (1987) "The Humean Theory of Motivation," *Mind*, New Series, Vol. 96, No. 381 (Jan., 1987) pp. 36-61.

Spinoza, Baruch. (1883) *Tractatus Politicus*, (ed. and intro.) R.H.M. Elwes, (trans.) A.H. Gosset (London: G. Bell &Son).
-------, (1930) *Ethics*, from *Spinoza Selections* (ed.) John Wild (Charles Scribner's Sons).

Stocker, M. (1979) "Desiring the Bad: An Essay in Moral Psychology," *Journal of Philosophy*, 76: 738-753.

Strawson, Galen. (1994) "The Impossibility of Moral Responsibility," *Philosophical Studies* 75, p. 5-24.
-------, (2003) "Mental Ballistics and Second Order Belief," *Proceedings of the Aristotelian Society* N.S., 103: 227-56.
-------, (2004) "Against Narrativity," *Ratio (new series)* XVII 4 December 2004, pp. 428-452.
-------, (2009) Selves: *An Essay in Revisionary Metaphysics* (Oxford: Clarendon Press).

Strawson, P.F. (1982) "Freedom and Resentment," in Gary Watson, (ed.), *Free Will* (Oxford: Oxford University Press).

Taylor, James Stacey, (ed.). (2005) *Personal Autonomy New Essays on Personal Autonomy and Its role in Contemporary Moral Philosophy* (Cambridge: Cambridge University Press).

Taylor, Shelley E. and Brown, Jonathon D. (1988) "Illusion and Well Being: A Social Psychological Perspective on Mental Health," *Psychological Bulletin*, Vol. 103, No. 2.

Tiberius, Valerie. (2008) *The Reflective Life: Living Well With our Limits* (Oxford: Oxford University Press).

Trungpa, Chogyam (1976) *The Myth of Freedom and the Way of Meditation*, (ed.) John Baker and Marvin Casper, foreword Pema Chodron (Boston: Shambhala Publications).

Twain, Mark. (1984) *The Adventures of Huckleberry Finn* (Penguin Classics: New York).

Van Inwagen, Peter. (1989) "When Is the Will Free?" *Philosophical Perspectives 3: Philosophy of Mind and Action Theory*: 399-422.

Velleman, David J. (1992) "What Happens When Someone Acts?" *Mind* 101:461-81.
------, (2000) *The Possibility of Practical Reason* (Oxford: Oxford University Press).

------, (2008) “The Way of the Wanton” from Watkins and Mackenzie (2008).

Vohra, Ashok. (1986) *Wittgenstein's Philosophy of Mind*, (London: Croom Helm).

Vohs, Kathleen D. and Schooler, Jonathan W. (2008), “Encouraging Belief in Determinism Increases Cheating,” *Psychological Science*, Vol. 19—Number 1, p. 49-54.

Watkins, Kim and MacKenzie, Catriona, (eds.). (2008) *Practical Identity and Narrative Agency* (Routledge).

Watson, Gary, (1975). “Free Agency” *The Journal of Philosophy*, Vol. LXXII, No. 8, April 24, 1975.

Wegner, Daniel M. (2002). *The Illusion of Conscious Will* (Cambridge, Massachusetts, London, England: Bradford Book, MIT Press).

Weiner, David Avraham, 1992. *Genius and Talent,* (Cranberry, NJ: Associated University Presses).

Williams, Bernard. (1981) *Moral Luck* (Cambridge: Cambridge University Press).
------, (1985) *Ethics and the Limits of Philosophy* (Cambridge: Harvard University Press).
Wittgenstein, Ludwig. (1961) *Notebooks*, *1914-1916* (ed.) V.H. Von Wright and G.E.M. Anscombe, (trans.) G.E.M. Anscombe (New York: Harper and Brothers).
-------, (1974) *Tractatus Logico-Philosophicus*, (trans.) D.F. Pears, and B.F. McGuiness, (intro.) Bertrand Russell (Routledge Classics, New York).
-------, (2001) *Philosophical Investigations*, (trans.) G.E.V. Anscombe (Blackwell Publishing Ltd.).

Zegzebski, Linda Trinkaus. (1996) *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*, (Cambridge University Press).

Zehr, Howard. (2015) *Changing Lenses: Restorative Justice For Our Times*, (Herald Press; 25th Anniversary Edition).

Zimmerman, David. (2002) “Reason-Responsiveness and Ownership-of-Agency: Fischer and Ravizza’s Historicist Theory of Responsibility,” *Journal of Ethics* Vol. 6, No. 3, pp. 199-234.